THREE ESSAYS ON DEVELOPMENT AND LABOR ECONOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yanan Li

May 2018

THREE ESSAYS ON DEVELOPMENT AND LABOR ECONOMICS

Yanan Li, Ph.D.

Cornell University 2018

My dissertation comprises three empirical studies on development and labor economics. The first two essays are in the context of National Employment Guarantee Schemes (thereafter, NREGA) in India. In the first one, I study whether existing inequality hinders the implementations of NREGA, using district-level data on land ownership distribution and the implementations of NREGA. To address potential endogeneity issues, I leverage a historical institution in India, the land revenue collection system established by British colonial rulers during 1750-1861, to construct an instrumental variable for land inequality. Both OLS and instrumental variable results give robust evidence that the concentration of land ownership reduces public works provision. This relationship could be explained by the mechanism that public works schemes raise agricultural wages in the private labor market, thereby incentivizing big landlords to use their political power to oppose this program. This paper provides the first empirical evidence that the concentration of landownership, a proxy for political power, is a hurdle to providing public employment to the poor. In the second essay, I focus on the participants of NREGA and study the wage bargaining effect of participating in NREGA. Using a household level panel and a difference-in-differences framework, I find indirect evidence that participating in NREGA would increase the wage bargaining power for both participants themselves and for their spouses in the private labor market.

The third essay focuses on a different social protection policy, minimum wage standards in China. I utilize a spatial lag methodology to study city-level strategic interactions in setting and enforcing minimum wage standards during 2004-2012 in China. This analysis finds strong evidence of spatial interdependence in minimum wage standards and enforcement among main cities in China. If other cities decrease minimum wage standards by 1 RMB, the host city will decrease its standard by about 0.7-3.2 RMB. If the violation rate in other cities increases by 1 percentage point, the host city will respond by an increase of roughly 0.4-1.0 percentage points. These interactions suggest the need for policy coordination in labor regulation.

BIOGRAPHICAL SKETCH

The author of the dissertation, Yanan Li, is a PhD candidate in the Dyson School of Applied Economics and Management at Cornell University. She received a bachelor's degree in agricultural economics and management from Zhejiang University in 2010 and a master's degree in agricultural economics from Renmin University of China in 2012. She will join the economics department in Beijing Normal University in 2018 fall.

To my grandmother, father, mother and sister.

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my special committee chair, Ravi Kanbur, and to the remaining members of my committee: Christopher Boone, Lawrence Kahn and Victoria Prowse. Thanks to Ravi for many insightful conversations and for guiding me through the PhD program. His passion in both research and teaching is always an inspiration. Thanks to Lawrence for always offering brilliant comments and suggestions on my projects. For teaching me a variety of empirical techniques, I am grateful to Victoria. Working with her has been one of the best experiences of my PhD by far. I am especially grateful to Chris for his constructive suggestions on my job market paper and his support during the job market season. Without the kind help and support of my committee, this dissertation would not have been possible.

During my PhD study, I also greatly benefited from interacting with other professors and researchers both in and outside of Cornell University. Special thanks go to Yanyan Liu and Shanjun Li, who have given me substantial help and support over the years. I am also grateful to Christ Barrett, Arnab Basu, Leah Bevis, Nancy Chau, Ritam Chaurey, John Hoddinott, Mallika Thomas, Xiaobo Zhang and Laura Zimmermann, who have each kindly offered very helpful comments on my dissertation.

One of the best things at Cornell is to be surrounded by talented peers. Special thanks go to Jakina Debnam, Naveen Sunder and Haoying Wang, who have offered extensive discussions on my dissertation and helped me through the hard times. I also want to thank Maulik Jagnani, Vesall Nourani, Tanvi Rao and the attendees from AEM development research seminar.

Winter in Ithaca is usually long. I feel extremely fortunate to have had the best cohort, including Jennifer Denno Cissé, Jakina Debnam, Xiaoli Fan, Jie Li and Linden McBride. I will remember forever those days when we were young and together. I have also received help and support from many others at Cornell including Ying Ji, Lisha Liu, Hang Lü, Jing Qian, Xiaowen Shi, Vidhya Soundararajan, Yue Wang, Kai Wu, Dafeng Xu, Yi Xu, Linlin Zhang, Yang Zhang and many others.

In addition, I would like to thank Jun Xiong, who is always there to listen and support. I would also like to thank my undergraduate professor Hongdong Guo, for his encouragement throughout the years. And thanks to my coauthor Wangyang Lai for his help in the job market season.

Lastly, I would like to thank my best friend, Wenjun Wang, for always being there. Thanks to my mother and father for giving me this incredible educational opportunity.

May 10, 2018

at Bogotá, Colombia

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

LAND INEQUALITY AND THE PROVISION OF PUBLIC WORKS—EVIDENCE FROM NATIONAL RURAL EMPLOYMENT GUARANTEE SCHEME IN INDIA

1.1 Abstract

Does existing inequality hinder redistributive policies that aim to help the poor? This paper answers this question under a widely used redistributive policy in developing countries—public works schemes. Using district-level data on land ownership distributions and the implementations of the National Rural Employment Guarantee Scheme in India, I find robust evidence that the concentration of land ownership reduces public works provision. This relationship could be explained by the mechanism that public works schemes raise agricultural wages in the private labor market, thereby incentivizing big landlords to use their political power to oppose this program. To address the potential endogeneity due to unobservables and measurement error, I leverage a historical institution in India, the land revenue collection system established by British colonial rulers during 1750-1861, to construct an instrumental variable for land inequality. Due to the concentration of post-independence land reforms enacted in landlord-dominated areas, those areas have lower land inequality today than the previously non-landlord dominated areas. The IV estimates suggest that a 1 percent increase of land Gini coefficient would lead to a 3-6 percent decrease in public job provision. The results are robust to using the alternative measurements of land inequality and public works implementation. To exclude the possibility that the higher provision of public jobs in more equal areas is driven

by a higher demand for public jobs, I show that more equal areas have higher agricultural wages in the private labor sector. This paper provides the first empirical evidence that the concentration of land ownership, a proxy for political power, is a hurdle to providing public employment to the poor, suggesting power asymmetries could hinder policies aimed at promoting equity.

1.2 Introduction

Does inequality lead to more or less redistributive efforts to the poor? This question has been studied extensively using theoretical models, with the earlier literature suggesting a positive association (e.g. Alesina and Rodrik, 1994; Persson and Tabellini, 1994) and the more recent literature suggesting a negative association (e.g. Benabou, 2000; Galor et al., 2009). The empirical evidence is relatively lacking in identifying the direction of the effect and the mechanisms through which inequality might affect redistributive policy, with a few exceptions (e.g. Boustan et al., 2013; Cinnirella and Hornung, 2016; Ramcharan, 2010).

The inequality of land ownership is an important form of inequality, as land is the main production factor before the industrial economy and still today in many developing countries. Furthermore, the distribution of land is directly linked to the concentration of political power. This power gravitates towards landlords, who may either influence tenants' votes or directly influence the politicians in the direction beneficial to themselves. The literature has provided evidence that large landlord elites influence the political process to prevent economic reforms or redistributive policies, such as educational expenditure (Cinnirella and Hornung, 2016; Ramcharan, 2010), human-capital accumulation (Galor et al., 2009), general social assistance programs (Anderson et al., 2015) and public goods (Beg, 2016).

In line with these recent studies, this paper answers the question regarding land inequality and redistribution under another widely used redistributive policy—public work schemes, which, due to its complexity in design and implementation, warrant special attention. A public works program is the provision of employment at a prescribed wage for those unable to find alternative employment by the creation of public infrastructure projects, such as transport infrastructure (e.g. roads, railroads and canals) and public services (e.g. sewage and dams). It is financed by the government and functions as a form of social safety net in many developing countries, such as India, Philippines, Bangladesh and Chile (Subbarao, 1997). The provision of public jobs raises agricultural wages in the private labor market (Imbert and Papp, 2015), thereby incentivizing big landlords to use their political power to oppose this program. India is the perfect context studying the relation between land inequality and the provision of public works, because it has the world's largest public works program—the National Rural Employment Guarantee Scheme (thereafter, NREGA) and faced with a historical tension arising from land inequality.

In this paper, I compare district (within-state) variations of land ownership inequality and public works provision, using census data on district-level land distribution in 2005 and the implementation data of the NREGA program since its inauguration in 2006. Land inequality is measured by the Gini coefficient. The provision level of public employment is measured by four dimensions: the fraction of rural households provided with employment, the per capita labor expenditure, average days of employment provided per person in either Schedule Caste or Schedule Tribe (thereafter, SC/ST) and the total number of completed works per rural person. OLS estimates suggest that a 1 percent difference in land Gini coefficient leads to a 0.6-1 percent gap in NREGA provision.

To address the potential endogeneity issue arising from measurement errors and omitted variables in the OLS estimation, I use a historical institution as the instrumental variable for land inequality—the land revenue collection system established by British colonial rulers during 1750-1861. This variable derives from the study by Banerjee and Iyer (2005). Despite a higher Gini coefficient of land ownership inequality in landlord-dominated areas during 1885-1948, such areas experienced more frequent land reforms after Indian independence. Therefore, the first-stage conditional correlations suggest that landlord-dominated districts have significantly lower Gini land ownership inequality in 2005. Under the assumption that the instrument is exogenous, the IV estimates confirm the negative effect of land ownership inequality on public works schemes. 2SLS estimates suggest that a 1% difference of land Gini coefficient leads to a 3-6 % gap in NREGA provision. To examine the sensitivity of the results to the exclusion condition, I construct bounds for the 2SLS estimates following Conley et al. (2012). The negative effect still holds when relaxing the exclusion restriction of the instrumental variable by allowing a negative association between the instrument variable and NREGA provision and a slight positive relation between these two variables.

Both OLS and IV results are robust when using the alternative measurement of land inequality—the share of land owned by the top 10% largest farmers, which more directly captures the top distribution and hence large farmers' political power. I finally exclude the possibility that the higher provision of public jobs in more equal areas is due to a higher demand for public jobs, by showing these more equal areas have higher agricultural wages in private sector.

To make sense of the results, compare two districts A and B with similar socio-economic characteristics. If district A's land Gini Coefficient is 1% (equivalent to 0.0047 in absolute terms) larger than that in district B, then district A will have 5% fewer households provided with NREGA jobs; per capita NREGA labor expenditure in district A will be 4% lower than that in district B; each person in Schedule Caste or Schedule Tribe in district A will on average work 3% fewer days than those from district B; the total number of works per rural person completed in district A will be 6% lower than that in district B.

This study adds to the understanding of the heterogeneity of the implementation of NREGA across different districts. NREGA claims to provide 100 days of working opportunity to each rural household in need of jobs. As a matter of fact, however, there is an un-met demand for jobs in almost all districts and the extent of un-met demand differs by districts. Existing literature has been trying to explain this heterogeneity of NREGA implementation mostly in terms of political incentives and administrative capacity (Gulzar and Pasquale, 2016; Niehaus and Sukhtankar, 2013; Nath, 2015; Gupta and Mukhopadhyay, 2016; Sheahan et al., 2016), and of the political reservation system (Dunning and Nilekani, 2013; Bose and Das, 2015). To the best of my knowledge, this paper is the first study to link district-level heterogeneity in the provision of NREGA jobs to the inequality of land ownership distribution. Districts with more concentrated land distributions are expected to see a lower provision of NREGA employment, because in those districts big farmers have stronger political power to block wage-increasing public works schemes. Indeed, there is abundant anecdotal evidence showing big farmers lobby to suspend the provision of NREGA employment (e.g. Maiorano, 2014), but broad-based quantitative testing of this notion has not been attempted previously. [1]

Investigating the question of land inequality and public works provision adds to the understanding of Indian land inequality which, as a legacy of British colonial institutions, has been a historically important and intricate issue. The relation between landlords and the landless affects different aspects of rural life and shapes the effectiveness of public policies. There has been a large number of land reforms since Indian independence, but most of them are through legislated ceilings on landholding (rather than direct land redistribution) and such reforms have been rarely implemented with any degree of seriousness (Besley and Burgess, 2000). As a result, after all those land reforms, the share of land occupied by the top 10% biggest farmers is still about 46%. This paper shows that the concentration of land ownership hence political power is a hurdle to redistributive efforts and successful anti-poverty policies, and offers a potential justification for further efforts at land reform. Moreover, compared to the estimates derived using soil or other geographical information as instrumental variables, the IV estimates in the current paper are particularly policy relevant because the lower levels of land inequality seem to be driven by land reforms (rather than natural conditions).

[1]In studying clientilism between landlords and the landless in Indian villages, Anderson et al. (2015) show land-owning elites will prefer weak provision of centrally funded pro-poor prgrams such as Employment Guarantee Program. The current paper differs from their paper in at least three respects. First, their survey data is restricted to 3 regions in the state of Maharashtra, while the current study uses district-wise nationally representative data. Second, they proxy landlords' political power by the proportions of land in the village dominated by the upper caste, Maratha. I use the concentration of land ownership, which goes beyond the constraint of caste backgrounds and have more general implications. Third, the pro-poor policy in their paper, EGS, is a previous form of NREGA. It is believed that NREGA has incorporated the lessons and successes of EGS, with broader goals and better implementations.

This paper speaks to the general discussion of inequality and public expenditures. The literature finds a detrimental effect of early inequality on the emergence of human-capital accumulating and growth-promoting institutions (e.g. Persson and Tabellini, 1994; Sokoloff and Engerman, 2000; Galor et al., 2009). The main mechanism is that land concentration induces landowners to use political power to assure lower public expenditure in education, for fear that higher public education investment would raise up labor cost or generate migration from agricultural sector to industrial sector. This mechanism also applies in the context of public works schemes. Providing public employment to the landless and the marginal farmers will increase labor wages (Imbert and Papp, 2015), and this wage effect will incentivize landlord elites to oppose the implementation of the public works schemes (Anderson et al., 2015; Maiorano, 2014).

This paper also broadly speaks to the literature on inequality, redistribution and economic growth. This literature initially argues that inequality is conducive to the adoption of growth-retarding redistributive policies (Alesina and Rodrik, 1994; Persson and Tabellini, 1994). Under democratic societies, if the median voter is poorer than the average voter, then the majority vote will lead to high tax rates and more redistribution to the poor, which impedes investment and economic growth. This mechanism is supported by some existing literature (Boustan et al., 2013). However, the current paper, coupled with other recent empirical evidence (e.g Galor et al., 2009; Ramcharan, 2010), casts doubt on this underlying mechanism. In contrast, the evidence suggests that inequality is a hurdle for redistribution, provided that the landlords, or better-endowed agents, have sufficient political power to influence redistribution policies.

The remainder of this paper is organized as follows. In section 2, I discuss the background information of the NREGA, highlighting the necessary facts that make it possible for landlords to play a role in the provision of NREGA jobs. Section 3 discusses the mechanism of how land inequality affects public works provision. Section 4 discusses data issues. Section 5 presents the empirical strategy and principal findings, followed by robustness checks relaxing the perfect exogeneity restriction and using alternative measurements of the NREGA implementation and land inequality. Section 7 concludes.

1.3 Background: The National Rural Employment Guarantee Act

1.3.1 Demand-Driven Nature of NREGA Employment

The Mahatma Gandhi National Rural Employment Guarantee Act of 2005 created the "right to work" for all households in rural India through the National Rural Employment Guarantee Scheme. It was a three-phased nation-wide rollout, with 199 districts in Phase 1 (Feb 2006), 128 districts in Phase 2 (April 2007) and the remaining 261 districts in Phase 3 (April 2008). By 2008, it reached all districts in India. It is the largest public works program in the world so far and asserts guaranteeing 100 days of working opportunity for each household per financial year (June in the current year to May in next year). Households need to obtain job cards from the local governments, which are used to record work done and payment. According to the Act, as long as an eligible household files applications for jobs, the local government must provide employment within

15 days and within 5 kilometers of the applicant's home. Otherwise, states are liable to pay unemployment allowances. However, in practice there are still frictions in the implementation leading to some unmet demand, such that those wanting work do not get it in a timely manner.

More than half of the works are related to water conservation, with other types of works including irrigation provision, land development and rural connectivity. Wages are to be paid at the statutory minimum wage rates, which makes this program a means of enforcing minimum wage laws. The wage rate is job specific rather than gender specific, as opposed to the private labor market where women earn a much lower wage rate than men. Therefore, NREGA jobs are especially appealing for women. As a social insurance tool, NREGA has a stronger demand in backward areas with poor agricultural conditions, such as bad soil and weather. For instance, Santangelo (2016) finds workers resort to NREGA to a larger extent when the local economy is hit by worse agricultural productivity shocks.

1.3.2 Financing NREGA and the Supply Constraint

The National Rural Employment Guarantee Act incentivises States to provide employment by stating that 100 percent of the unskilled labor cost and 75 percent of the material cost of the program is borne by the central government. The labor to material ratio could vary from 90:10 to 60:40.

The overall annual labor spending on NREGA at state/district/ block/ village level is a pre-determined cap. Labor budget for each financial year is determined in the previous year, following a "bottom-up" process from the village level to the state level and last to the central government (NREGA Operational Guidelines, 2013). This budget plan includes (i) the anticipated quantity of demand for jobs in the next year (ii) the precise timing of the demand for work and (iii) a shelf of projects to be prepared and prioritized to meet job demand. Table 1.1 presents the various steps involved in the preparation and finalization of annual labor budgets. Because labor budget is an estimation and NREGA is a demand driven program, the Act states that the States may, based on actual performance, any time during the year, come back to the Ministry requesting revision of their existing labor budget, following the procedures in Table 1.1. However, in fact, the flexibility is limited. Once the labor budget is finalized, the maximum supply of jobs in each state/district/block will not be changed for the next financial year.

Therefore, there will be a shortage of supply for NREGA jobs if any of the following cases occurs—(i) an exogenously fixed maximum level of spending on NREGA by the center government; (ii) an underestimation of job demand in the budget planning; (iii) a poor timing of job demand; and other cases. The actual implementation is further complicated by states' constraints in organizing projects and workers. Even if the budget planning is not an issue, accommodating supply to demand could still be a challenge because of the incapability to meet the relatively skilled labor requirements at the local level, such as panchayat technical assistants (Dutta et al., 2014). As a result, although the NREGA program is designed to be a demand-driven program, there is an un-met demand for jobs in almost all states (Dutta et al., 2014). On average, each household works roughly 35 person-days per financial year, far less than the claimed 100 days. The extent of the un-met demand differs by districts and by time.

Table 1.1: Timelines for various steps involved in preparation and finalization of annual labor budget.

Date	Action to be taken
15th August	Gram Sabha to approve GP Annual Plan and submit to PO
15th September	PO submits consolidated GP Plans to Block Panchayat
2nd October	Block Panchayat to approve the Block Annual Plan and submit to DPC
15th November	DPC to present District Annual Plan and LB to District Panchayat
1st December	District Panchayat to approve District Annual Plan
15th December	DPC to ensure that shelf of projects for each GP is ready
31st December	Labour Budget is submitted to Central Govt.
January	Ministry scrutinizes the Labour Budget and requests for compliance for deficiencies, if any
February	Meetings of Empowered Committee are held and LB finalized
February, March	Agreed to LB communicated to States. States to feed data of Month wise and District wise breakup of "Agreed to" LB in MIS and communicate the same to Districts/ blocks GPs
Before 7th April	States to communicate OB, Center to release upfront / 1st Tranche.

Source: Mahatma Gandhi National Rural Employment Guarantee Act 2005 – Operational Guidelines, 4th version. Chapter 6.10.

1.3.3 Landlords and NREGA Employment

In addition to supply constraints in implementing the NREGA, landlords could also affect the supply of public jobs. Providing public jobs to the landless and marginal farmers will increase labor wages (Imbert and Papp, 2015), which will potentially increase production costs for landlords who hire casual labors. Thus this wage effect brings landlords an economic incentive to oppose the implementation of NREGA (Anderson et al., 2015; Maiorano, 2014). There are at least two stages where big landlords can intervene the process of providing NREGA jobs.

First, at the stage of making the labor budget, landlords may lobby against a budget plan that provides enough jobs to the rural poor. As Table 1.1 shows, budget planning is a bottom-top decision making process. The demand for

Table 1.2: Summary statistics of NREGA implementation

	(1) 2006 mean/sd	(2) 2007 mean/sd	(3) 2008 mean/sd	(4) 2009 mean/sd	(5) 2010 mean/sd
% of households provided employment	12.40	17.69	27.32	29.63	31.09
	(23.01)	(25.16)	(26.11)	(24.22)	(23.81)
avg days of employment provided per rural SCST person	2.01	2.95	5.24	6.57	6.45
	(4.55)	(5.07)	(6.60)	(6.93)	(7.12)
avg days of employment provided per rural woman	1.25	1.79	3.48	4.32	4.11
	(3.66)	(3.97)	(5.50)	(5.90)	(5.45)
avg days of employment provided per NREGA-woman	19.00	16.35	17.00	21.61	20.84
	(16.27)	(17.03)	(16.80)	(15.88)	(13.09)
labor expense per rural person (2006 Rs.)	68.64	92.41	163.12	175.46	167.70
	(192.87)	(179.99)	(258.53)	(224.40)	(202.98)
number of completed works per 1000 rural persons	1.73	5.06	9.48	13.39	7.68
	(5.06)	(17.57)	(26.09)	(20.69)	(13.50)
# of districts with employemnt provided	122	202	409	410	415
Observations	416	416	416	416	416

Notes: Original data come from MGNREGA public portal. Only districts used in regression analysis are included. Labor expense is deflated by state-wise Consumer Price Index, using 2006 as the base year.

NREGA jobs and the shelf of projects are first identified at the Gram Panchayat level, then the demand and supply are consolidated at the block level, and further aggregated at the district and state levels. The fact that lower level governments such as block and village have a substantial discretion in this process renders big landlords' influences very likely. It is after all easier for landlords to lobby village governments than state governments.

Second, even after labor budget is made, big farmers can still use their political power to block the implementation, such as delaying work assignment, payment and some complementary machinery (see Maiorano (2014) for anecdotal evidence of lobbying). As a result, as NREGA annual report shows, the final work completed is smaller than the original budget.

1.4 Mechanism

The political mechanism of inequality and redistribution has been established by the literature. Higher inequality lowers the level of awareness of the poor, decreasing the level of their political participation (e.g. Bardhan and Mookherjee, 2005; Ramcharan, 2010). Meanwhile, greater inequality can concentrate the benefits of political participation and simplify the collective action problem among the landed, which leads to a higher and more effective political participation among the landed elites. In the cases that the landlord elites are a net loser from redistribution, they would block redistribution. Therefore, a higher land inequality predicts lower redistributions to the poor.

As the primary interest of this paper lies in economic effects rather than political effects, I will impose a crude political mechanism under which landlords have sufficient political power against redistributive policies. Instead, I will focus on the economic incentives that lead big landlords to oppose the provision of public employment.

Providing public jobs to the poor introduces a competition for labor between the public works schemes and the rural private employers. The literature has found that the introduction of NREGA increases rural casual labor wages by 6 percent (Imbert and Papp, 2015). This wage effect could potentially reduce landlords' profit, if they keep hiring casual labor. Therefore, the wage-increasing nature of public works schemes provides big landlords the economic incentives to oppose the program.

1.5 Data

1.5.1 Land Inequality

District-wise data on land distribution in 2005 come from Indian Agricultural Census (excluding Maharashtra), which is conducted at five yearly intervals. Although the information is collected on operational land holdings rather than owned land holdings, the wholly owned and self-operated holdings accounted for 97.14 percent (Page 29, Agriculture Census Report, 2005). Therefore, I use this dataset on operational land holdings to approximate the distributions of land ownership in India. [2]

This dataset has information on the number and area of operational holdings across the following size bins (in 1000 hectares): below 0.5; 0.5-1; 1-2; 2-3; 3-4; 4-5; 5-7.5; 7.5-10; 10-20; 20 & above. [3] I use the average size of land holdings in each bin to construct land ownership Gini coefficient. The first row of Table 1.3 shows, the average Gini coefficient in our sample districts is 0.47. Figure 1.1 provides state-wise average Gini coefficients in 2005. Figure 1.2 provides the

[2]According to Agriculture Census in India, "an Operational holder is the person who has the responsibility for the operation of the agricultural holding and who exercises the technical initiative and is responsible for its operation." An operational unit could include multiple plots. The operated areas comprise of i) Land owned and self operated; ii) Land leased in; iii) Land otherwise operated.

[3]I use the information on "Sub-total" land holdings, including both individual holding and joint holdings, to measure district level land distribution. The ratio of joint holdings to individual holdings is, 1:6.5 in terms of numbers and 1:5 in terms of areas (Agriculture census report 2005, page 121). Land operated by institutions constituted less than 0.5% of the total area, and is excluded from the data.

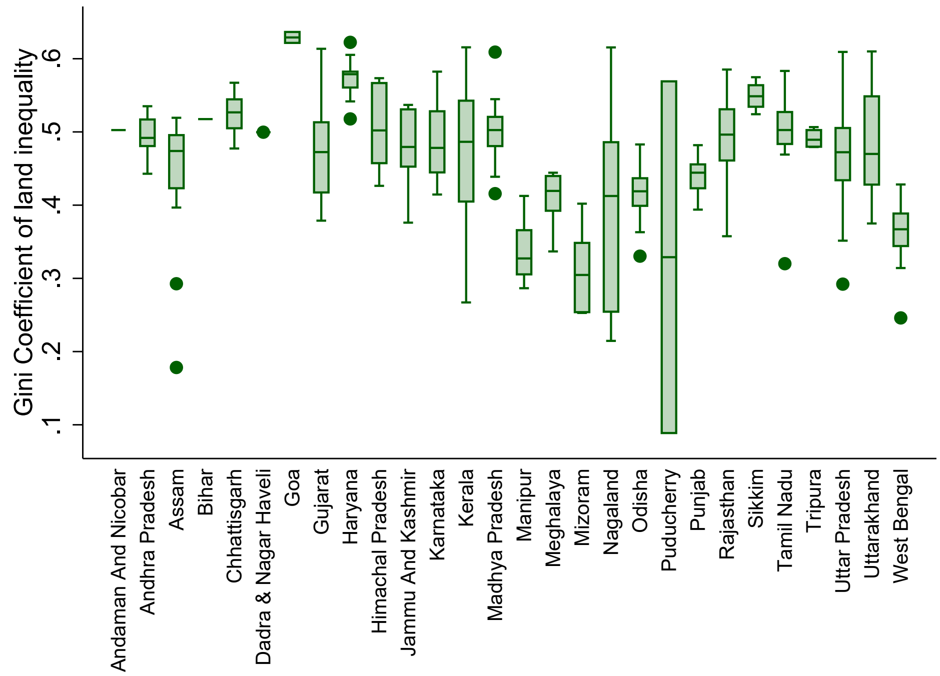


Figure 1.1: Land inequality (Gini coefficient) by state, 2005

Source: The author calculated Gini coefficient based on district-wise land distribution data from 2005 Indian Agricultural census. Only states in the OLS regression sample are included.

shares of operated areas by each decile of holdings. The largest 10 percent of operational holders operate about 46 percent of total land in India. Figure A.2 in Appendix plots the probability distributions of Gini coefficients. [4]

[4]There is another relevant fact that supports the legitimacy of using the concentration of land holdings to proxy for landlords' political power. According to Agriculture Census Report, 2005 (Page 34), about 96.0 percent of the operational holdings and 94.7 percent of operated areas were operated by village residents whose entire area of land holdings was locating in the village of his residence. These high ratios reduce the concern that big landlords, whose land is in the village but who themselves live outside of the village, might not have the political power to affect NREGA in the current village.

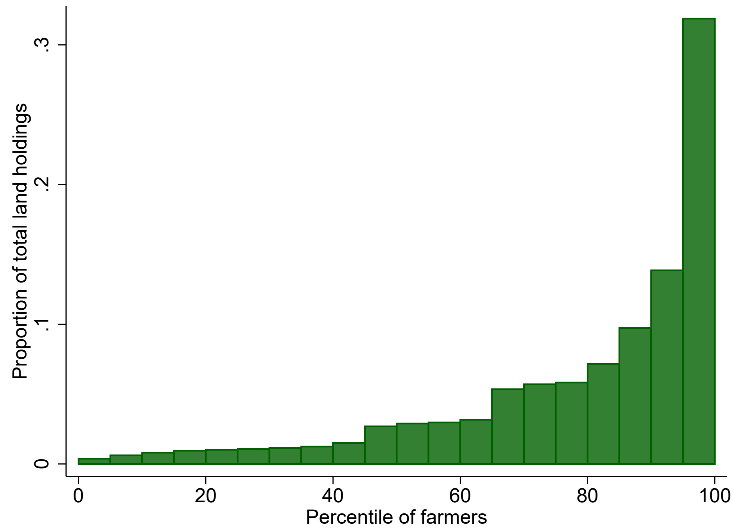


Figure 1.2: Shares of land area by percentiles of holdings, 2005

Note: Size-classes are as follows: below 0.5; 0.5-1; 1-2; 2-3; 3-4; 4-5; 5-7.5; 7.5-10; 10-20; 20 & above. The graph is derived by first ranking all land holdings by class size in India, then calculate the share of land operated at each decile.

1.5.2 NREGA Implementation

NREGA implementation data come from public data portal[5]. Table 1.2 presents summary statistics of NREGA implementation by financial year (starting from April in the current year and ends in March the next year) using alternative measurements. The first row tells that, among all working population in India, 12% of them worked for at least one day in public works in 2006, the first year that NREGA was introduced. This number increased to 18% in 2007, and 30% in 2010.

Labor expenses are deflated by state-level consumer price index, using 2006 as the base year. The average wages per rural person received (regardless of their work status in NREGA) increased from 68 Rupees in 2006 to 167 Rupees

[5] $mnregaweb4.nic.in/netnrega/dynamic2/dynamicreport\_new4.aspx$

in 2010. Figure A.1 in Appendix plots probability distributions of per capita labor expenditure. When focusing only on the subpopulation that were provided public employment, the average wages that each household received increased from 2667 Rupees in 2006 to 2867 Rupees in 2010.

Information on the three-phased roll-out comes from the document by NREGA Report (2007) [6]. Phase 1 include 200 districts, phase 2 includes 130 districts and phase 3 includes the rest of districts. Phases are determined based on the ranking of Backwardness Index (Zimmermann, 2012). I extract this index and its five components from Indian Planning Commission 2003 Report, including agricultural wages in 1996, agricultural productivity per person 1990-93, agricultural productivity per hectare 1990-93, ratio of SC/ST in the population from 1991 census and poverty ratio 1994 (Commission et al., 2003).

I motivate this paper by observing substantial heterogeneity of NREGA implementations across districts. Therefore, it's important to show variations of both NREGA and land inequality across districts. Figure 1.1 and Figure 1.3 present state level variations of public works provision and land inequality. Figure 1.4 visually presents a negative relation between land inequality and public works provision by doing a kernel regression of the shares of households provided with public jobs on the share of land occupied by top 10% biggest farms.

1.5.3 Demographic and Geographic Information

District profiles are downloaded from 2001 population census, including caste composition, employment and industry structure, literacy rate, amenities and

[6] This online document nicely presents the phase-in progress $http://nrega.nic.in/MNREGA\_Dist.pdf$

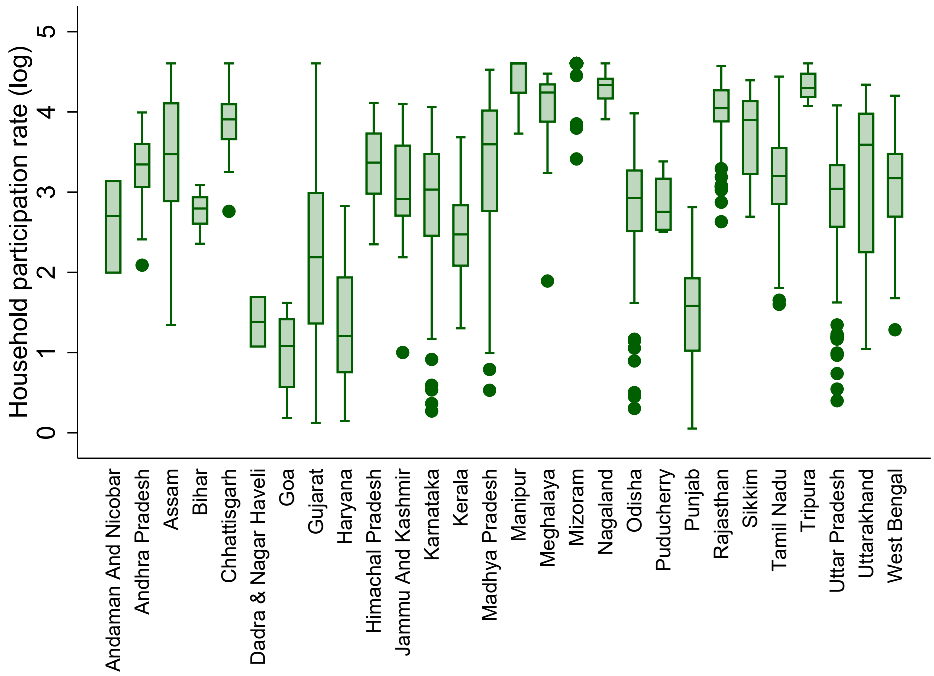


Figure 1.3: Shares of Households provided with NREGA employment by state, 2006-2010

Note: Shares are calculated as total number of households provided with NREGA employment divided by total rural households in the district. All districts in the OLS regression sample are included.

infrastructural facilities, district area size and so on. Population between 2001 and 2010 are filled using these two years' census data, assuming a growth rate equal to that during 1991-2001. Table 1.3 presents summary statistics of district-wise demographic information in 2005.

The monthly rainfall data are obtained from Center for Climatic Research, University of Delaware. Indian agricultural year is split into two distinct seasons– wet season (from June to November) and dry season (from December to May). Existing studies document that NREGA participation is strongly associated with rainfall shocks in wet season. Therefore, I compute wet sea-

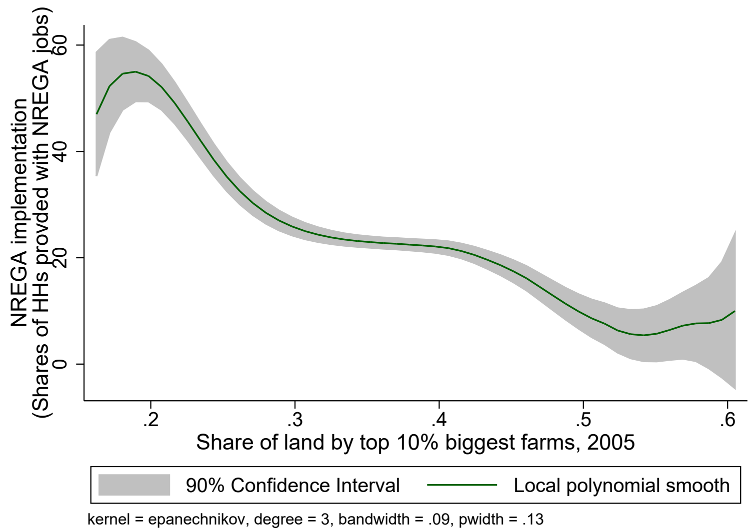


Figure 1.4: Local polynomial smoothing

Note: Author calculates the share of land by top 10% biggest farms based on 2005 India Agricultural census. NREGA implementation is measured by the share of households provided with public jobs. Kernel = epanechnikov, degree=3, bandwidth=.09.

son precipitation by aggregating the amount of precipitation between June and November in the study year. Figure A.4 in Appendix presents annual rainfall deviation in the country. Soil information is obtained from Food and Agriculture Organization (FAO) Digital Soil Map of the World and Derived Soil Properties (CDROM). I map geographical coordinates to district level soil texture. Table 1.3 shows that 91% of the land contains medium or fine level soil; 9% of land is covered by course soil.

Compiling these data sets into a district-wise panel is complicated by district jurisdictional changes during 2001-2011. There were 640 districts in 2011, as opposed to 593 districts in 2001 (Census, 2011). In the analysis, districts with boundary changes are excluded, although results are robust to adding these

Table 1.3: Descriptives of land inequality and demographic information

	mean	sd	min	max
Gini coef.	0.47	(0.08)	0	1
Rural area (Sq. km)	5044.34	(4862.42)	119	45382
wet season rainfall (100 mm)	1.03	(0.69)	0	6
% of land covered in fine soil	20.08	(24.19)	0	97
% of land covered in medium soil	70.70	(27.68)	0	100
% of Rural pop	77.94	(15.42)	12	100
Literacy rate	65.29	(11.70)	31	97
% of SC population	15.84	(8.84)	0	50
% of ST population	15.89	(26.18)	0	98
Work-population ratio	40.91	(6.98)	24	63
% of Main workers	30.68	(5.95)	17	52
% of Marginal workers	10.23	(4.20)	2	24
% of Agricultural labourers	22.63	(12.98)	1	63
% of Cultivators	37.77	(18.04)	1	82
% of Household industry workers	4.05	(3.89)	1	31
% of Other industries	35.60	(17.71)	8	91
% villages with Safe Drinking water	96.19	(10.49)	24	100
% villages with Electricity (Power Supply)	84.76	(18.89)	10	100
% villages with Paved approach road	60.87	(25.51)	12	100
% villages with Primary school	84.35	(14.15)	31	100
% villages with Medical facility	41.82	(25.66)	3	100
% villages with Post and telephone facility	52.43	(26.81)	4	100
Observations	416			

Total workers = main workers + marginal workers = Ag laborers + cultivators + household industry workers + Other workers. Main workers were those engaged in any economically productive activity for 183 days or more during the year. Marginal workers were those who worked for less than 183 days. A person was considered as cultivator if he or she was engaged either as employer, single worker or family worker in cultivation of land owned or held from government (or private persons, institutions). A person was regarded as an agricultural labourer if she/he worked in another person's land for wages in cash, kind or share.

districts back. The final sample includes 416 districts at 2001 district level.[7]

1.6 Empirical Model and Results

1.6.1 OLS Results

I examine the effect of land inequality on public works provision by pooling the NREGA implementation data during 2006-2010 and using across district (within-state) variations of land concentration in 2005.[8] The model specification is:

$$Y_{it} = \alpha_0 + \beta * INE_{i,2005} + \alpha X_{it} + \alpha_s D_{state} + \alpha_t D_t + \varepsilon_{it}, \forall t \in \{2006, 2007, 2008, 2009, 2010\} \tag{1.1}$$

where β is the coefficient of interest.

$INE_{i,2005}$ denotes land inequality in district i in 2005, measured by Gini coefficient (in logarithm). Y_{it} denotes the implementation of the NREGA program in district i in year t, measured by proportions of rural households provided with NREGA employment (in logarithm). A negative sign of β means NREGA job provision is negatively associated with land inequality. Standard errors are clustered at the district level.

[7] This seemingly much smaller number relevant to total Indian districts is a consequence of combining different data sources, each of which has some missing districts. First, 2005 Agriculture Census only contains 528 districts (as Maharashtra state is not included), about 90 of which are either newly created or the original districts that got split. These 90 districts are dropped when I combine Agriculture census to 2001 population census. Second, 26 districts that are included in Agriculture census are not included in NREGA public portal (and this public portal posts information at 2010 district level). Third, about 10 districts lack information on soil quality or rural area size etc that further reduces the number of observations. In the end, we have 416 districts for each year at 2001 district level.

[8] Table A.2 shows that land distribution didn't change at statistically significant level during 2005 and 2010.

To identify the effect of the concentration of land ownership on public works provision, β, I need to control for variables that are correlated with land inequality and at the same time affect NREGA implementation. The first set of confounding factors contains the capacity of local governments to accommodate job supply to job demand. As mentioned in the previous section, although central government bears most of the cost, it does not imply there will be zero cost to local governments when employing workers under NREGA. The implementation cost may be particularly high in poor districts (Dutta et al., 2014). Therefore, I control for variables reflecting the level of local economic development to capture local governments' accommodation capability, such as the percentage of villages that have access to drinking water, electricity, paved road and schools in the district and other rural infrastructure variables.

In the same vein, I also control for "Backwardness Index", a score constructed by Indian planning commission in 2003, with smaller numbers meaning being more backward. The literature has shown that NREGA program rolls out from backward districts to more affluent districts, in the order of their rankings on this index(Zimmermann, 2012; Dasgupta et al., 2017). However, this roll-out rule is not absolutely enforced, reflected by the fact that many Phase 2 districts have smaller values of backwardness index than Phase 1 districts. Therefore, I also include phase dummies to capture the heterogeneous implementation by phases.

In addition, the model also includes soil texture and the current wet season's rainfall deviations from historical means, because these geographic variables could affect both the demand for and the supply of NREGA jobs, and are also documented to be associated with land distribution. [9] I also include a vec-

[9]This relation between geographic and climate information and land ownership distribution

tor of state dummy, D_{state}, and year dummy, D_t, restricting the cross-sectional comparisons to within-state variations.

The results of OLS estimates are presented in Table 1.4. It gives a significantly negative relation between Gini coefficient and the proportion of rural households provided with NREGA employment. The results are robust to adding extra covariates. Column 4 suggests, districts with a 1% (or in absolute term, 0.0047) higher Gini coefficient would have 0.6% (or in absolute term, 0.006 * 30=1.8 percentage points) fewer households provided with NREGA jobs.

1.6.2 Addressing Endogeneity

OLS estimates cannot be interpreted as causal. The most apparent threat is reverse causality, as redistributive policies such as public works schemes also shape inequality. Using land inequality in 2005 (much ahead of the initiation of NREGA) allows for some control of potential reverse causality (i.e. it's reasonable that land inequality in 2005 will affect public work provision in post-2006, but unlikely that public works in post-2006 will affect land inequality in 2005). However, the model will still capture some endogeneity issues if there are some omitted variables which are correlated with both land ownership and the demand/supply side of NREGA implementation. For instance, adverse geographical and climatic characteristics may concentrate land ownership by re-

is established by existing studies that use various geographical conditions to instrument for land inequality, including climatic information, soil quality and the share of cash crop (inequality-rising) and wheat/rice crop etc. (e.g. Easterly, 2007; Sokoloff and Engerman, 2000; Galor et al., 2009; Ramcharan, 2010; Cinnirella and Hornung, 2016; Baten and Juif, 2014). The spirits of these IVs are, small farmers are usually less able to hedge against negative weather shocks, and have a smaller demand for land in areas with poor soil quality (or in areas with violent rainfall variability). Thus, regions with poorer soil quality (or more rain variability) have higher land concentration.

Table 1.4: Dep var: % of households provided with NREGA jobs (OLS)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Gini coef. (log)	-0.642**	-0.645**	-0.632**	-0.638**
	(0.292)	(0.273)	(0.289)	(0.273)
log rural area(Sq. km)		0.291***	0.328***	0.293***
		(0.097)	(0.098)	(0.097)
log Rural population		-0.243***	-0.204**	-0.241***
		(0.083)	(0.089)	(0.083)
Literacy rate		-1.040***	-1.097***	-1.017***
		(0.345)	(0.390)	(0.344)
Wet season rainfall deviation		-0.092**	-0.094**	-0.092**
		(0.036)	(0.036)	(0.036)
% of land covered in fine soil		-0.272	-0.519**	-0.287
		(0.208)	(0.222)	(0.209)
% of land covered in medium soil		-0.375**	-0.556***	-0.402**
		(0.155)	(0.175)	(0.155)
% of Agricultural labourers		1.708***	2.087***	1.733***
		(0.456)	(0.467)	(0.457)
% of Main workers		-0.206	0.099	-0.228
		(0.748)	(0.839)	(0.745)
% of Marginal workers		1.202	1.826**	1.255
		(0.784)	(0.918)	(0.791)
% of SCST population		0.850***	1.472***	0.927***
		(0.264)	(0.263)	(0.266)
% villages with Safe Drinking water		1.365	1.880	1.448
		(1.209)	(1.263)	(1.220)
% villages with Electricity (Power Supply)		0.565**	0.512*	0.528*
		(0.280)	(0.289)	(0.282)
% villages with Paved approach road		-0.609**	-0.841***	-0.629**
		(0.277)	(0.295)	(0.278)
% villages with Primary school		0.084	0.185	0.080
		(0.328)	(0.343)	(0.329)
% villages with Medical facility		-0.505	-0.605	-0.501
		(0.398)	(0.401)	(0.400)
% villages with Post and telephone facility		-0.366	-0.390	-0.365
		(0.247)	(0.261)	(0.249)
Phase 2 indicator		-0.179**		-0.182**
		(0.071)		(0.071)
Phase 3 indicator		-0.519***		-0.526***
		(0.082)		(0.082)
Composite Backwardness Index			0.041	0.100
			(0.095)	(0.090)
State Dummies	Yes	Yes	Yes	Yes
Observations	1224	1224	1224	1224
R square	0.45	0.67	0.65	0.67

Notes: Dependent variable is the logarithm of the share of households provided with NREGA jobs in the district. Standard errors are in parentheses, clustered at district level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

ducing the demand for land by marginal farmers[9] ; and meanwhile adverse geographical and climatic characteristics may increase the demand for NREGA jobs and decrease the government capability to supply jobs. Although I have included the covariates of soil texture and rainfall variations, if there are other such geographical and climatic variables omitted, OLS estimates will be biased (upward and toward zero in the given example). In addition to omitted variable bias, there is also measurement error of land distribution, which will lead to the attenuation bias of the estimates.

I address the endogeneity issues by taking advantage of historical institutions in India — land revenue collection system, established by British colonial rulers during 1750-1861. This variable is constructed based on the study by Banerjee and Iyer (2005). Land revenue, or land tax, was the major source of government revenue in India and during British times as well. British administration established three systems to collect land revenue in all cultivable land in British India: (a) landlord-based system, where the liability for a village or a group of villages lay with with a single landlord; (b) an individual cultivator-based system, where revenue settlements was made directly with individual cultivators; (c) village-based system, where village bodies which jointly owned the village were responsible for the land revenue. Figure A.3 in the appendix presents the map of these three different land revenue systems in British India. System (c), village-based system, could be further grouped as either system (a) or (b), depending on whether the village body was a single landlord or a large number of members with each person being responsible for a fixed share of the revenue. Table 1.5 presents state-wise distribution of landlord and non-landlord districts.

Table 1.5: State-wise distribution of landlord and non-landlord districts

	Landlord	Non-landlord	Total districts	Mean landlord proportion
Andhra Pradesh	2	8	10	0.34
Bihar	1	0	1	1
Chhattisgarh	4	1	5	0.80
Gujarat	0	6	6	0
Haryana	0	4	4	0.15
Karnataka	0	11	11	0
Madhya Pradesh	10	1	11	0.89
Odisha	6	2	8	0.68
Punjab	0	5	5	0.14
Rajasthan	1	0	1	1
Tamil Nadu	2	9	11	0.28
Uttar Pradesh	12	34	46	0.59
Uttarakhand	0	3	3	0.38
West Bengal	10	0	10	1
Total	48	84	132	0.51

Source: This is a subsample of districts from the study by Banerjee and Iyer (2005).

To identify a causal relation between land distribution and the provision of public works under NREGA, I use the binary indicator of land revenue system — whether this district was a landlord district in British India — to instrument for land inequality in 2005. The instrumental variable strategy relies on the assumption that land revenue collection system under British India only affects redistributive policies through contemporary and current land inequality, after controlling for all observables. This is plausible because the way that British colonial rulers decided land revenue system in different areas was not based on

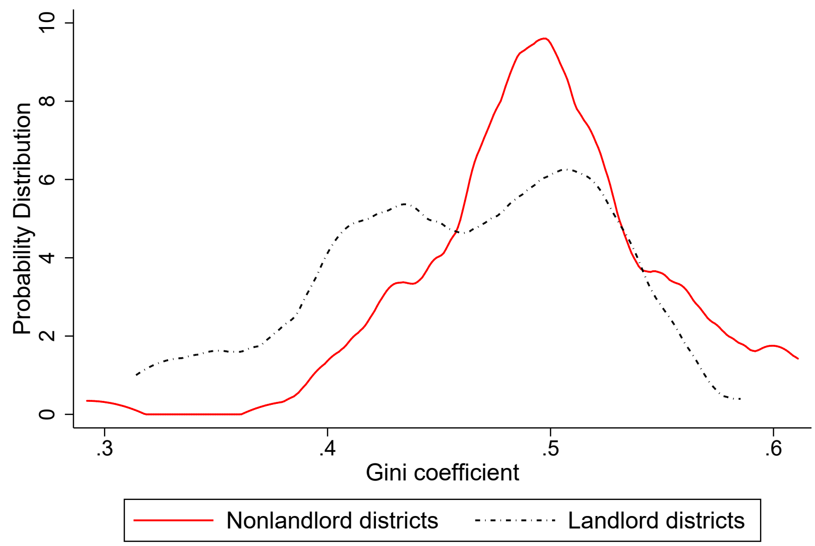


Figure 1.5: Visualize first stage —Land Inequality (Gini coefficient) in landlord/non-landlord districts

Source: The author calculated land ownership Gini coefficient based on 2005 India Agricultural census.

a hard rule, i.e. in terms of land fertility, weather or labor productivity (Banerjee and Iyer, 2005). Figure 1.5 visually presents the negative relation between landlord-dominated revenue collection system and current land inequality.

I estimate first stage relation using the following equation:

$$INE_{i,2005} = \alpha_0' + \rho Z_i + \alpha' X_{it} + \alpha_s' D_{state} + \alpha_t' D_t + \eta_{it} \tag{1.2}$$

where Z_i is the binary indicator that equals to 1 if district i used to be a landlord-dominated district in British India, and zero otherwise; $INE_{i,2005}$ denotes land inequality in district i in 2005, measured by Gini coefficient; X_i denotes the same vector of district-wise covariates as in Equation(1.1).[10]

[10]By restricting the variations to be within-state, in this IV estimation, states where land tenure systems don't vary across districts within the state will be absorbed in the state fixed effects, such as Bihar, Gujarat, Haryana, Karnataka, Punjab, Rajasthan, Uttarakhand and West Bengal. As a result, 91 districts in 6 states are left and contribute to the variations in the IV estimation.

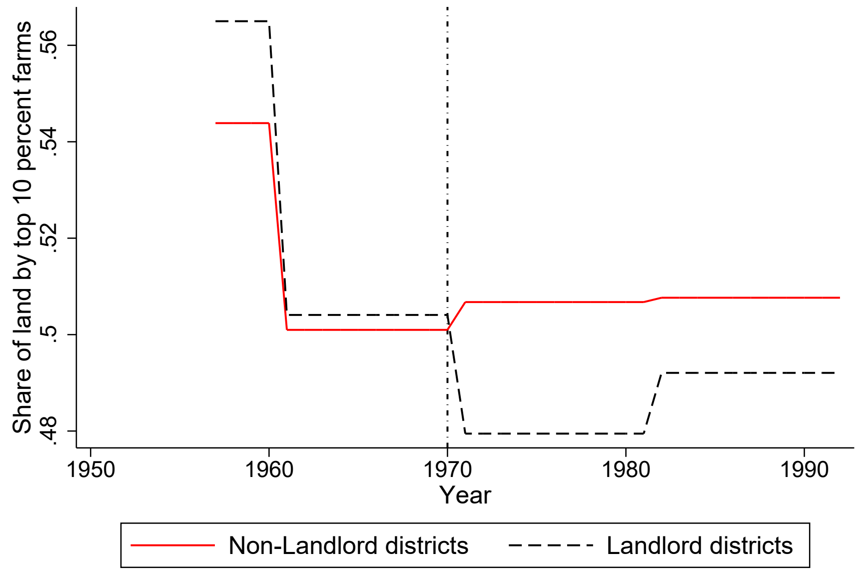


Figure 1.6: Trends of Land Inequality (share of land by top 10% land holdings) in landlord/nonlandlord districts

Source: Land reform data during 1957-1992 are from (Besley and Burgess, 2000). Data on Land revenue system are from (Banerjee and Iyer, 2005).

The first-stage conditional correlations suggest that landlord-dominated districts have 7.4% lower Gini land ownership inequality in 2005 (Table 1.6). This estimated effect is equivalent to -0.035 (= -7.4% * 0.47) in absolute term of gini coefficient, or 0.5 (=0.035/0.08) standard deviation of gini coefficient, considering that the mean and standard deviation of Gini coefficient are respectively 0.47 and 0.08 in the sample.

The first-stage result that previously landlord-dominated districts in British India has a lower land inequality today is consistent with the study by Banerjee and Iyer (2005). They show that states with a higher landlord proportion had higher Gini measures of land ownership inequality in 1885, and this inequality persisted until the end of the colonial period[11]. However, as they argue, major landlord-dominated states enacted 6.5 land reforms in the period between 1957-1992, while non-landlord states had an average of 3.5. According to Besley and Burgess (2000), states that enacted a larger number of land reforms had a greater decline of Gini coefficient of land inequality. Therefore, with this chain of reasoning, landlord district saw a greater decline of land inequality than non-landlord districts, driven by a great number of land reforms after Indian Independence. Furthermore, the negative sign of first-stage results is consistent with the study by (Besley et al., 2016) that shows in the long-run land inequality is lower in areas that saw greater intensity of tenancy reform.

I then plot the numbers of land reforms over time in major landlord and non-landlord states in Figure 1.7, which provides consistent evidence with the literature that landlord areas enacted more frequent land reforms than non-landlord areas.[12] To depict *when* land ownership distribution in landlord dominated districts started to become more equal than non-landlord dominated districts, I further plot the trends of land inequality, measured by the share of land owned by the top 10% land holdings, for major landlord and non-landlord districts in Figure 1.6.[13] It shows that the shift of landlord districts from having relatively high land inequality to relatively low land inequality occurred in 1970. Inter-

[11] Banerjee and Iyer (2005) explains why the choice of landlord revenue system had a strong effect on the distribution of land and wealth in British India period. "Under landlord-based systems, the landlords were given a more or less free hand to set the terms for the tenants and, as a result, they were in a position to appropriate most of the gains in productivity."

[12]I calculate these two series of numbers by combining state-wise land reform data from (Besley and Burgess, 2000) into the current district-wise sample. Major landlord-dominated areas are states with an above-median share of districts belonging to landlord dominated districts.

[13]Ideally, I need district-wise land inequality to plot the changes of land inequality over time in landlord versus non-landlord districts. However, the lack of district-wise land distribution data prevents me from doing so. Instead, by combining with state-wise data on land reform and land distribution from Besley and Burgess (2000), I plot these trends for major landlord states and non-landlord states, where landlord states are defined as states with an above-median share of districts belonging to landlord dominated districts, and non-landlord states.

estingly, this turning point coincides with the time when major landlord states started to outnumber non-landlord states in land reforms as shown in the Figure 1.7. All such information together explains the negative sign of first-stage estimate—landlord districts, although starting with higher land inequality in British Indian period, enacted more land reforms after Indian Independence, and hence ended up having lower land inequality in 2005.

The instrumental variable strategy relies on the assumption that land revenue collection system under British India only affects redistributive policies through contemporary and current land inequality, after controlling for observables. However, if different historical property rights institutions lead to persistent unobserved culture and institutional outcomes, and such unobserved outcomes are also correlated with redistributive policies, then this IV would violate the exclusion condition. I will examine the sensitivity of the estimations to the degree in which the exclusion restriction is potentially violated using sensitivity analysis proposed in Conley et al. (2012).

1.6.3 IV Results

Table 1.7 presents two-stage least square (2SLS) estimations of the effect of land inequality on NREGA implementation (measured by proportions of household provided with NREGA jobs). The sample size drops to one third of the original size, because the instrumental variable, landlord versus non-landlord dominated districts indicator, is only defined in districts that were under British India during 1850-1947.

Restricting the sample to the IV sample, column 1 shows OLS results with

Table 1.6: Dep var: Land inequality (gini coefficient) in 2005, (First stage)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Landlord district indicator	-0.079***	-0.082***	-0.074***	-0.074***
	(0.021)	(0.020)	(0.020)	(0.020)
log rural area(Sq. km)		0.037	0.043	0.041
		(0.031)	(0.030)	(0.028)
log Rural population		-0.010	-0.021	-0.020
		(0.040)	(0.036)	(0.037)
Literacy rate		-0.013	-0.002	-0.006
		(0.125)	(0.108)	(0.122)
Wet season rainfall deviation		-0.007**	-0.007**	-0.007**
		(0.003)	(0.003)	(0.003)
% of land covered in fine soil		-0.478***	-0.490***	-0.495***
		(0.073)	(0.067)	(0.072)
% of land covered in medium soil		-0.243***	-0.259***	-0.263***
		(0.059)	(0.056)	(0.057)
% of Agricultural labourers		0.144	0.178*	0.181*
		(0.101)	(0.097)	(0.100)
% of Main workers		0.558**	0.485**	0.489**
		(0.231)	(0.209)	(0.225)
% of Marginal workers		0.081	0.122	0.100
		(0.296)	(0.261)	(0.296)
% of SCST population		-0.110	-0.065	-0.063
		(0.114)	(0.106)	(0.116)
% villages with Safe Drinking water		-0.433	-0.408	-0.368
		(0.410)	(0.388)	(0.416)
% villages with Electricity (Power Supply)		-0.031	-0.046	-0.047
		(0.092)	(0.089)	(0.092)
% villages with Paved approach road		0.171**	0.161**	0.165**
		(0.083)	(0.077)	(0.080)
% villages with Primary school		0.127	0.117	0.112
		(0.092)	(0.085)	(0.085)
% villages with Medical facility		0.011	0.009	0.007
		(0.057)	(0.059)	(0.058)
% villages with Post and telephone facility		0.001	0.003	0.004
		(0.070)	(0.068)	(0.067)
Phase 2 indicator		-0.002		-0.005
		(0.021)		(0.021)
Phase 3 indicator		0.010		0.000
		(0.023)		(0.022)
Backwardness Index			0.067***	0.066***
			(0.018)	(0.017)
State Dummies	Yes	Yes	Yes	Yes
Observations	570	570	570	570
R square	0.63	0.80	0.81	0.81
F test: landlord indicator coef=0	14.21	17.13	14.12	14.28

Notes: "Landlord district indicator" equals 1 if the district in question was a landlord district (i.e. landlords were responsible for collecting land revenue) in British Raj. Land Gini coefficient is constructed using 2005 Indian Agricultural census. All models include year dummy. Standard errors are clustered at district level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

a full set of control variables, as a reference to 2SLS estimates. It suggests that a 1% increase in gini coefficient of land ownership is associated with a 1.7% (or equivalent to 0.5 percentage points, given that the average share of households provided with NREGA jobs is 30%) decrease of the share of households provided with NREGA jobs. Column 2-5 present IV estimates with different sets of covariates added. First-stage F statistics are all above 10, suggesting a rejection of weak instrument null hypothesis. The size of the point estimate for the Gini coefficient is relatively stable over the last four IV specifications, suggesting that a 1% increase in gini coefficient of land ownership would have decreased the share of households provided with NREGA jobs by 5% (or equivalent to 1.5 percentage points, given that the average share of households provided with NREGA jobs is 30%).

Column 3 in Table 1.7 adds various variables that reflect the demand for and the supply of NREGA jobs. It suggests soil quality and some rural infrastructure are negatively associated with public works provision. For instance, a 1% higher proportion of land covered by fine soil rather than coarse soil in the district is associated with a 2% lower participation by households. Similarly, a 1% higher proportion of villages having access to post and telephone facility is associated with a 1% lower participation by households. Such evidence is consistent the demand-driven nature of NREGA — areas with favorable agricultural conditions may not need the job protection by NREGA, because there are already enough jobs in rural labor market. The negative relation may also be because the central government allocates less public employment in those relatively better endowed and more developed areas.

Column 3 also contains two Phase indicators, using Phase 1 as the reference

Table 1.7: Dep var: % of households provided with NREGA jobs

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
Gini coef. (log)	-1.711***	-4.892**	-4.558***	-4.844***	-5.124***
	(0.557)	(1.916)	(1.508)	(1.647)	(1.741)
log rural area(Sq. km)	0.626***		0.711***	0.862***	0.767***
	(0.232)		(0.241)	(0.209)	(0.238)
log Rural population	-0.429**		-0.509***	-0.444**	-0.577***
	(0.170)		(0.178)	(0.179)	(0.182)
Literacy rate	-0.433		-0.384	0.381	-0.349
	(0.653)		(0.739)	(0.750)	(0.780)
Wet season rainfall deviation	-0.067		-0.091	-0.092	-0.092
	(0.062)		(0.061)	(0.063)	(0.063)
% of land covered in fine soil	-0.939*		-2.130***	-2.311***	-2.517***
	(0.521)		(0.715)	(0.772)	(0.823)
% of land covered in medium soil	-0.563*		-1.250***	-1.352***	-1.512***
	(0.340)		(0.456)	(0.483)	(0.508)
% of Agricultural labourers	2.393***		2.924***	3.681***	3.270***
	(0.820)		(0.913)	(1.016)	(1.012)
% of Main workers	-0.292		1.303	2.572	1.193
	(1.802)		(1.969)	(2.035)	(2.061)
% of Marginal workers	-1.493		-1.299	1.070	-1.207
	(1.692)		(1.910)	(1.991)	(2.068)
% of SCST population	0.339		-0.205	0.536	-0.014
	(0.519)		(0.550)	(0.647)	(0.586)
% villages with Safe Drinking water	1.035		-0.333	0.226	-0.291
	(3.443)		(3.497)	(3.640)	(3.611)
% villages with Electricity (Power Supply)	0.546		0.601	0.247	0.461
	(0.424)		(0.556)	(0.579)	(0.576)
% villages with Paved approach road	-0.877*		-0.257	-0.593	-0.208
	(0.450)		(0.539)	(0.544)	(0.562)
% villages with Primary school	1.194***		1.521***	1.739***	1.517***
	(0.432)		(0.436)	(0.449)	(0.448)
% villages with Medical facility	-0.564		-0.547	-0.556	-0.563
	(0.464)		(0.451)	(0.487)	(0.446)
% villages with Post and telephone facility	-1.236***		-1.291***	-1.475***	-1.267***
	(0.413)		(0.388)	(0.406)	(0.397)
Phase 2 indicator	-0.342***		-0.351**		-0.370***
	(0.124)		(0.137)		(0.141)
Phase 3 indicator	-0.541***		-0.487***		-0.547***
	(0.145)		(0.150)		(0.159)
Composite Backwardness Index	0.138			0.292	0.419**
	(0.165)			(0.209)	(0.178)
State Dummies	Yes	Yes	Yes	Yes	Yes
Observations	457	457	457	457	457
First-stage F statistics		14.26	17.58	14.65	15.02

Notes: Column 1 shows OLS results; Column 2-5 present IV estimates with different sets of covariates added. Dependent variable is proportions of household provided with NREGA jobs in each year 2006-2010. District-wise land Gini coefficient is constructed using 2005 Indian Agricultural census. Instrumental variable is a binary indicator that equals 1 if the district in question was a landlord district (i.e. landlords were responsible for collecting land revenue) in British Raj. All models include year and state dummy. Standard errors are clustered at district level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

group. The estimated coefficients of Phase 2 and Phase 3 indicator are both negative, with the former having a smaller magnitude than the latter. This suggests that Phase 1 districts have the highest level of public employment provided, followed by Phase 2 districts and then Phase 3 districts. This relative position resonates with the fact that NREGA rolls out from the most backward districts to richer districts.

Column 4 replaces Phase dummies with "Backwardness Index", which captures how the program was rolled out. Column 5 adds both Phase dummies and Backwardness Index. Backwardness Index (a greater value indicating economically more developed) is expected to be negatively related to job provision, because more NREGA jobs are demanded in backward areas. However, both column 4 and column 5 suggest a positive relation between this index and NREGA job provision, although not at a statistically significant level. This positive relation may be because the higher demand in the backward areas is dominated by the lower capacity to accommodate NREGA projects. In addition, the magnitude of the estimated effect of land inequality on NREGA provision is slightly bigger in column 5 than that in column 3.

To put these results into perspective, consider the difference between land inequality Gini coefficient in two districts in Uttar Pradesh, Ballia and Allahabad. In Ballia, gini coefficient of land ownership was 0.486 (which is at the 50th percentile of the distribution of gini coefficient) in 2005, and in Allahabad, this number is 0.519 (which is at the 80th percentile). Using the estimates in column (5), the difference of 0.033 points, or 6.6% (= 0.033/0.518), in gini coefficient implies that 33% (=6.6 * 5 %) more households would have been provided with NREGA jobs in Allahabad if it had a land gini coefficient as small as Bellia's.

Considering the shares of households provided with NREGA jobs are, respectively, 22 percent in Bellia and 12 percent in Allahabad, this 33% increase would have eliminated one third (=33% * 12 / (22-12)) of the actual gap in job allocation rates between these two districts.

Both OLS and 2SLS estimations suggest a negative relation between land inequality and NREGA provision. In terms of the magnitude, 2SLS coefficient is about 3.8 times the OLS coefficient, suggesting that OLS results are biased upward (toward finding zero effect). A simple and possible source of endogeneity that leads to the upward bias of OLS results is measurement error in land ownership distributions. As I approximate the size of land by the average size of land holdings in the size bin it belongs to, it unavoidably creates noise. Another source of bias might be some omitted variables that lead to less job provision in more equal areas. The three-times difference between IV estimates and OLS estimates is also in line with other studies that use geographical conditions to instrument for land inequality (Easterly, 2007; Cinnirella and Hornung, 2016; Ramcharan, 2010).

It's noted that I take the logarithm of the dependent variable and gini coefficient of land ownership so that the estimated effect can be easily interpreted as percent changes. This is especially convenient when we compare estimates across different measurements of NREGA implementation in the robustness check section. The disadvantage of taking logarithm lies in losing more than 100 observations that have zero NREGA jobs provided, most of which are Phase 2 and Phase 3 districts before 2007. Therefore, to address the concern that the estimated results may be driven by sample selections, I use level regressions which will include districts that have zero NREGA jobs provided as well. The estimates are provided in Table A.3 in the Appendix. The level of dependent variable and gini coefficient here are standardized by the standard deviation of the sample observations. Column 5 suggests that a 1 standard deviation increase of gini coefficient is associated with 0.7 standard deviation decline of the share of households provided with NREGA jobs.

1.7 Robustness of the Identification

1.7.1 Robustness to Violations of Perfect Exogeneity

The credibility of 2SLS estimations rests on the identification assumption that the historical institution of landlord versus non-landlord dominated areas does not directly relate to the provision of public works other than through land distribution. However, this instrument variable may only be plausibly exogenous rather than perfectly exogenous. To gain a sense about the sensitivity of the estimated effect to the relaxation of perfect exogeneity conditions, I examine the bounds that we can place on the true effect of land inequality on public works provision, following the method proposed by Conley et al. (2012). This method has been used in other studies to examine the sensitivity of estimation results to the violations of exogeneity conditions (e.g. Nunn and Wantchekon, 2011; Ding et al., 2009).

This method relaxes IV exclusion restriction by allowing the instrument variable to also enter linearly in the second-stage regression with a coefficient γ [14].

[14] The following equation is a generalization of this method proposed by Conley et al. (2012).

$$Y = \beta X + \gamma Z + \varepsilon$$

Conley et al. (2012) show how to obtain the bounds for the IV estimate of the effect of interest (in the current paper, the effect of land inequality on public works provision, β) with prior information or assumptions about γ.

Applying the "Union of intervals" approach proposed by Conley et al. (2012) to the current paper, I find that if $\gamma < 0$, the bounds of β are actually further away from zero relative to the IV estimate of β which assumes perfect exogeneity (i.e. $\gamma = 0$). In other words, if landlord-dominated areas are still associated with less NREGA employment even after controlling for all covariates (which is likely if we regard public work projects as one kind of public investments and follow the argument by Banerjee and Iyer (2005) that landlord areas have lower public expenditures today), then IV estimates provide an underestimation (in terms of absolute value) of the true effect of land inequality on NREGA job provision.

Applying the same "Union of intervals" approach, I find that if $\gamma > 0$, the bounds of β will be closer toward zero relative to the IV estimate of β, and therefore IV estimates provide an overestimation (in terms of the magnitude) of the true effect. In other words, if landlord-dominated areas are still associated with more public employment even after controlling for all covariates, then IV estimates provide an overestimation (in terms of absolute value) of the true effect of land inequality on NREGA job provision. A such example is that the relatively backward landlord areas have a higher demand for public jobs for reasons not captured by the model. Figure 1.8 plots 95% confidence intervals for an array of assumptions about prior information of γ — the support of γ is assumed to be

where γ reflects how close the exclusion restriction is satisfied. Without prior information or assumptions about γ, the parameters β and γ can not be jointly identified. The IV exclusion restriction is equivalent to the prior belief that $\gamma = 0$. The definition of plausible exogeneity is having prior information that implies γ is near 0 but perhaps not exactly 0.

$[0, \delta]$. Because no distribution about the prior information is assumed in "Union of intervals" approach, this approach provides a conservative estimate of the bounds. For the 95% confidence interval of β to include zero, γ must be above 0.2. Now the question is, how likely will γ be above 0.2? I was trying to estimate γ by extracting a subsample which has the same level of land inequality and then regressing Y on Z and all other covariates. However, the small sample nature of the data set in this paper doesn't support this estimation.

To gain a deeper insight on the direction of the potential bias (the sign of γ), I compare economic development variables in landlord-dominated versus non-landlord-dominated districts, as shown in Appendix Table A.1. It shows that landlord-dominated areas are more backward today, with smaller proportions of villages having access to safe drinking water, electricity, paved road, medical facility, post and telephone facility, having lower labor productivity and agricultural wages. The underdevelopment in landlord districts could potentially not only lead to a higher demand for public jobs, but also a lower local capability to supply public jobs. Therefore, I cannot conclude from this comparison in which direction landlord areas are associated with NREGA job provision (i.e. the sign of γ is still unclear), which is especially because the empirical model already controls for all these observed economic development variables.

What I can conclude from this robustness check is, the true effect of land inequality on NREGA job provision will be stronger if landlord areas are directly associated with less NREGA provision through other mechanisms than land distribution. In the opposite, the true effect will be smaller if landlord areas are associated with more NREGA provision through other mechanisms than land distribution.

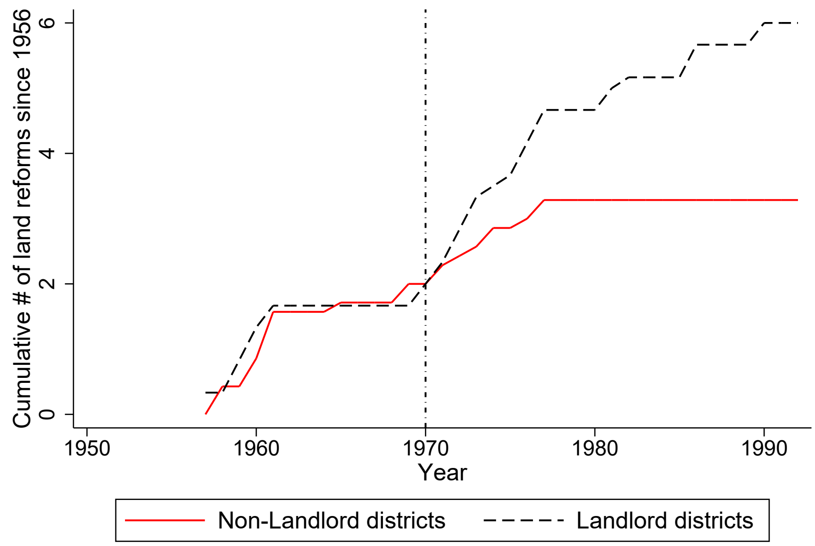


Figure 1.7: Frequencies of land reforms

Source: Besley and Burgess (2000) provides information on the cumulative number of land reforms at state level. This figure plots reform frequencies by whether this state has an above-median or below-median share of districts that used to belong to "landlord-dominated" areas in British India.

1.7.2 Alternative Measurements of NREGA Implementations

Both the OLS and IV results presented in the previous section use proportions of households provided with public employment to measure NREGA implementations. Table 1.8 presents estimates using alternative measurements of public works implementation — respectively, per capita labor expenditure, average days of employment provided per person in either Schedule Caste or Schedule Tribe, and the total number of completed works per rural person.

Panel A shows OLS results using the full sample. Panel B and C present OLS estimates and IV estimates using the IV sample. The instrumental variable is still the binary indicator that equals 1 if the district used to be a landlord district

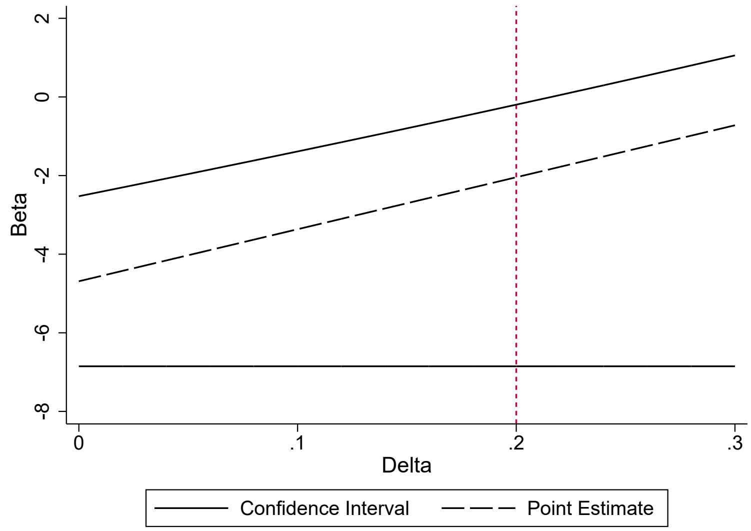


Figure 1.8: Plausibly exogenous bounds: Union of Intervals approach

This figure presents 95% confidence intervals for the effect of land inequality on shares of households provided with public works using unions of intervals approach (Conley et al., 2012). The confidence interval imposes the prior information that the support of γ, the sign on the instrument, is the interval $\gamma \sim [0, \delta]$.

in British Raj. The sample size drops in the IV sample because I have historical data on land tenure only for British districts. OLS results using the full sample show 0.4-1 percent decrease in districts with a 1 percent higher Gini coefficient. OLS results in the IV sample show a slightly higher effect, around 1 percent. 2SLS results give an even higher estimate for Gini coefficient, between 3-6 percents. The results using these alternative measurements are overall consistent with the estimates using household participation rate in NREGA.

Table 1.8: Robustness checks: Alternative measurements of NREGA implementation

	(1) Labor Expenditure	(2) Persondays	(3) # of projects
Panel A: OLS using the full sample			
Gini coef. (log)	-0.523*	-0.476*	-0.992***
	(0.288)	(0.276)	(0.375)
Observations	1223	1179	1199
R square	0.81	0.66	0.45
Panel B: OLS using the IV sample			
Gini coef. (log)	-1.172*	-1.181*	-1.137
	(0.685)	(0.704)	(1.109)
Observations	456	449	452
R square	0.83	0.68	0.45
Panel C: 2SLS			
Gini coef. (log)	-3.638*	-3.177*	-5.789
	(1.946)	(1.798)	(3.623)
Observations	456	449	452
Kleibergen-Paap rk Wald F statistics	15.03	14.85	14.17

Notes: The table shows district-level OLS and IV estimates using three alternative measurements of NREGA implementation. The dependent variables in each model are, respectively, (log) per capita labor expenditure, (log) average days that each person in Schedule Caste or Schedule Tribe worked in NREGA, (log) total number of completed works per rural person. Panel A shows OLS results using the full sample. Panel B shows OLS results using the IV sample. Panel C shows 2SLS estimates, where the instrumental variable is the binary indicator that equals 1 if the district used to be a landlord district (i.e. landlords were responsible for collecting land revenue collection) in British Raj. All specifications include a full set of covariates, including year and state fixed effects, phase indicators, backwardness index and other covariates listed in the last column of Table 1.7. Standard errors are clustered at the district level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

1.7.3 Does Gini Coefficient Capture the Top Distribution?—Alternative Measurement of Land Inequality

As Gini coefficients of land ownership reflect the whole distribution of land holdings, a natural question arises—does gini coefficient capture the top distribution of land ownership? It is after all the top, rather than the middle or bottom, distribution of land holdings that reflects the concentration of big farms and big landlords' political power. Therefore, I construct shares of land owned by the top 10% largest land holdings to measure land inequality, following Besley and Burgess (2000). As mentioned in previous section, land distribution data has information on the number and area of operational holdings across the following size bins (in 1000 hectares): below 0.5; 0.5-1; 1-2; 2-3; 3-4; 4-5; 5-7.5; 7.5-10; 10-20; 20 & above. I use the average size of land holdings in each bin to proximate land size in each group, then calculate shares of operated area for each decile of holdings. Figure 1.2 shows that the largest 10 percent of landlords own about 46 percent of the land area nationwide.

The scatter plot of Gini coefficient and the shares of land owned by the top 10% land holdings in Figure 1.9 shows that these two measurements of land inequality have the same trends. The shares of land owned by the top 10% land holdings are higher wherever Gini coefficients are greater. This provides descriptive support that differences in Gini coefficients between districts are able to capture the relative differences of large landlords' land holdings. Then I use this alternative inequality measurement of land ownership distribution to re-estimate the effect of land inequality on public works provision.

Table 1.9 presents the first-stage results, indicating landlord districts have a lower land inequality today. Table 1.10 present district-level OLS and IV estimates using the alternative measurement of land inequality, each model using an alternative measurement of NREGA implementation—share of households participating in NREGA employment, per capita labor expenditure, average days that each person in either Schedule Caste or Schedule Tribe worked in NREGA, the total number of completed works per rural person. Panel A shows OLS results using the full sample. Panel B and C shows OLS and 2SLS results using the IV sample, where the instrumental variable is the historical land tenure indicator. The estimations all give negative signs on land inequality. For instance, Panel C in column 1 shows when the share of land owned by the top 10% land holdings increases by 1 percent (or in the absolute term, by 0.46 percentage points, given the fact that the average shares of land owned by the top 10% is 46 percent), the shares of households provided with NREGA jobs will decrease by 6 percent (or equivalent to 0.06 * 30%=1.8 percentage points, given that the average share of households participation is 30%). In addition, similar to using Gini coefficient as the measurement of land inequality, 2SLS estimates have greater magnitudes than OLS estimates.

1.7.4 Is NREGA Demand Higher in More Equal Areas?

This paper mainly argues, those more equal areas have more public jobs provided to the rural poor because of the less interference of landlords, rather than through other means such as a higher demand for jobs. The previous sections conclude this argument by controlling for a series of demand side factors in the empirical model such as Backwardness Index and the percentage of agricultural labor. However, it is still important to examine the possibility of a higher

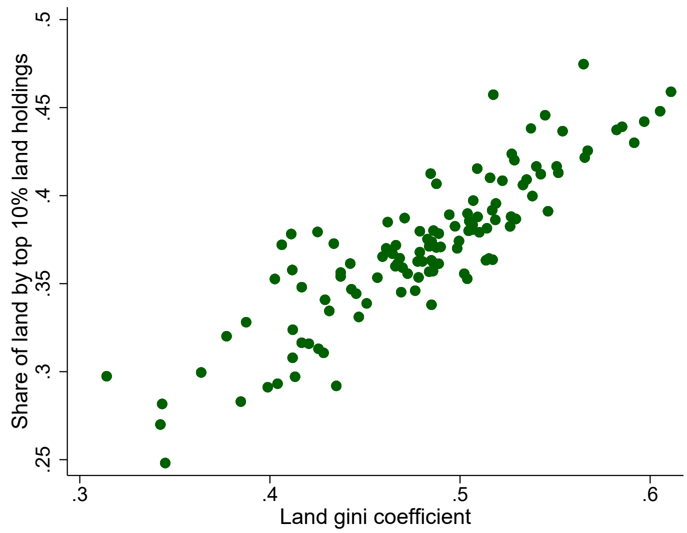


Figure 1.9: What does Gini coefficient capture? — Gini coefficient VS top 10% land holdings

Note: 2005 India Agricultural census.

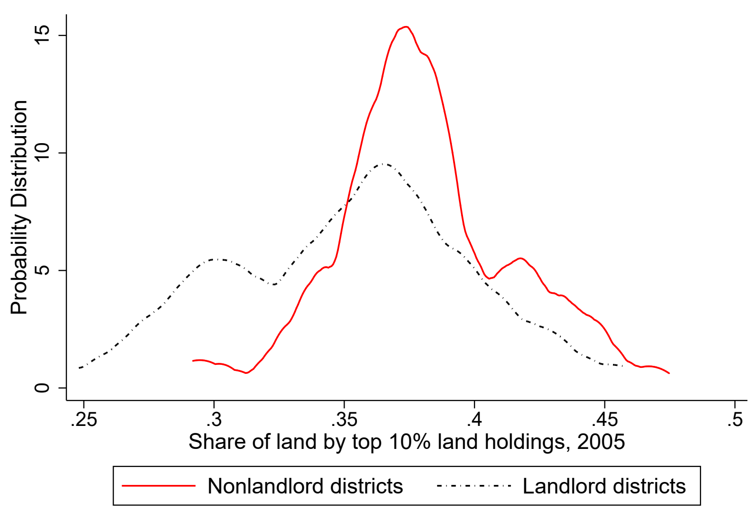


Figure 1.10: Visualize first stage — Land Inequality (top 10% land holdings) in landlord/non-landlord districts

Note: The proportion of overall land owned by the top 10% biggest land holdings in the district is constructed using 2005 Indian Agricultural census.

Table 1.9: Robustness check: Dependent variable—Share of land by top 10% land holdings (First Stage)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS
Landlord district indicator	-0.062***	-0.068***	-0.061***	-0.060***
	(0.017)	(0.018)	(0.017)	(0.017)
log rural area(Sq. km)		-0.015	-0.011	-0.009
		(0.032)	(0.032)	(0.030)
log Rural population		0.050	0.048	0.039
		(0.039)	(0.035)	(0.036)
Literacy rate		0.200*	0.216**	0.207*
		(0.107)	(0.089)	(0.105)
Wet season rainfall deviation		-0.005	-0.004	-0.004
		(0.003)	(0.003)	(0.003)
% of land covered in fine soil		-0.366***	-0.402***	-0.384***
		(0.076)	(0.074)	(0.075)
% of land covered in medium soil		-0.212***	-0.243***	-0.230***
		(0.065)	(0.065)	(0.062)
% of Agricultural labourers		0.025	0.095	0.064
		(0.089)	(0.087)	(0.086)
% of Main workers		0.399	0.387*	0.318
		(0.246)	(0.220)	(0.241)
% of Marginal workers		0.430	0.445*	0.443
		(0.303)	(0.247)	(0.304)
% of SCST population		-0.112	-0.040	-0.068
		(0.099)	(0.095)	(0.100)
% villages with Safe Drinking water		-0.463	-0.241	-0.418
		(0.481)	(0.483)	(0.496)
% villages with Electricity (Power Supply)		-0.167**	-0.201***	-0.186**
		(0.076)	(0.074)	(0.076)
% villages with Paved approach road		0.150**	0.143**	0.142**
		(0.070)	(0.062)	(0.068)
% villages with Primary school		-0.022	-0.045	-0.035
		(0.093)	(0.091)	(0.087)
% villages with Medical facility		-0.024	-0.034	-0.026
		(0.059)	(0.058)	(0.059)
% villages with Post and telephone facility		0.050	0.054	0.054
		(0.063)	(0.065)	(0.062)
Phase 2 indicator		0.007		0.004
		(0.022)		(0.022)
Phase 3 indicator		-0.012		-0.023
		(0.021)		(0.022)
Composite Backwardness Index			0.063***	0.069***
			(0.017)	(0.018)
State Dummies	Yes	Yes	Yes	Yes
Observations	457	457	457	457
R square	0.64	0.77	0.79	0.79
F test: landlord indicator coef=0	13.97	14.39	12.63	12.15

Notes: Dependent variable is the proportion of land owned by the top 10% biggest land holdings, constructed using 2005 Indian Agricultural census. "Landlord district indicator" equals 1 if the district in question was a landlord district (i.e. landlords were responsible for collecting land revenue) in British Raj. All models include year and state fixed effects. Standard errors are clustered at the district level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

Table 1.10: Robustness check: land inequality measured as the share of land by top 10% land holdings

	(1)	(2)	(3)	(4)
	% HHs participate	Labor Expenditure	Persondays	# of projects
Panel A: OLS using the full sample				
Share of land by top 10% (log)	-0.595**	-0.533*	-0.397	-0.692*
	(0.281)	(0.311)	(0.274)	(0.404)
Observations	1224	1223	1179	1199
R square	0.67	0.65	0.66	0.45
Panel B: OLS using the IV sample				
Share of land by top 10% (log)	-1.435**	-1.253*	-0.930	-1.121
	(0.567)	(0.643)	(0.652)	(1.099)
Observations	457	456	449	452
R square	0.62	0.66	0.68	0.45
Panel C: 2SLS				
Share of land by top 10% (log)	-6.529***	-4.673**	-4.076*	-7.439*
	(2.241)	(2.290)	(2.242)	(4.440)
Observations	457	456	449	452
Kleibergen-Paap rk Wald F statistics	12.15	12.23	12.12	11.80

The table shows district-level OLS and IV estimates using the alternative measurement of land inequality, and using four alternative measurements of NREGA implementation. District-wise land inequality is measured by "Share of land by top 10%" — the proportion of overall land owned by top 10% biggest land holdings, constructed from 2005 Indian Agricultural census. The dependent variables in each model are, respectively, (log) share of households participating in NREGA employment, (log) per capita labor expenditure, (log) average days that each SC/ST person worked in NREGA, (log) total number of completed works per rural person.

Panel A shows OLS results using the full sample. Panel B shows OLS results using the IV sample. Panel C shows 2SLS estimates, where the instrumental variable is the binary indicator that equals 1 if the district used to be a landlord district (i.e. landlords were responsible for collecting land revenue collection) in British Raj.

All specifications include a full set of covariates, including year and state fixed effects, phase indicators, backwardness index and other covariates listed in the last column of Table 1.7. Standard errors are clustered at the district level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

demand for public jobs in areas with more equal land distributions, which will shed light upon what is driving the smaller job provision in these areas. As the sample size in the 2SLS estimation drops by two thirds compared to the original OLS sample, I examine the relation between gini coefficient and economic development indicators separately for these two samples, and then explore how agricultural wages bias the estimated effect for each model.

The top panel in Table 1.11 shows the relation between gini coefficient and economic development in the full sample. By some measures of economic indicators, such as the fractions of villages with access to electricity, paved road and schooling, areas with more equal land distributions (i.e. lower Gini coefficients) are less developed. Backwardness Index and agricultural productivity (Rupees per hectare), however, are about the same in areas with different levels of Gini coefficients. In particular, agricultural wages, an important labor market variable, are slightly higher in these more equal areas. With higher agricultural wages in private sectors, the demand for NREGA jobs is presumably lower in such areas, hence eliminating the concern that the higher participation rate in these areas are caused by job demand rather than job supply. Column 1 and 2 in Table 1.12 reaffirm this argument by showing that adding the additional covariate, agricultural wages, into the original OLS model does not change the estimated effect much, and in fact, it slightly raises the estimate.

The second panel in Table 1.11 shows that in the IV sample, other than literacy rates and access to schools, economic characteristics do not vary by Gini coefficient at 10% significant level. The fact that labor market characteristics are not worse in areas with more equal land distributions teases out the possibility that the higher NREGA participation in these equal areas is driven by a higher demand for public jobs. The comparisons in Column 3 and 4 (and Column 5 and 6) in Table 1.12 further reaffirm this argument by showing that adding the additional covariate, agricultural wages, into the old OLS model (IV model) only slightly affects the estimated effects.

The examination of the relation between Gini coefficient and economic development also provides insight on the advantage and disadvantage of OLS and IV estimates. As mentioned earlier, OLS estimates have the potential endogeneity issue arising from unobserved geographic and climate variables. By switching to IV estimation, I can address the endogeneity issue in this regard, but meanwhile introduce another issue due to the change of sample representativeness. When sample size drops to 1/3, some properties of the original full sample disappear. For instance, as column 1 in Table 1.11 shows, the more equal areas have higher agricultural wages in the full sample but not in the IV sample. As higher agricultural wages in local markets indicate a lower demand for NREGA jobs, the difference of this property in these two samples probably explains why OLS estimates have smaller magnitudes than 2SLS estimates.

1.8 Conclusion

This paper studies how the concentration of land ownership, a proxy for landlords' political power, affects the effective implementation of public works schemes in the context of The National Rural Employment Guarantee Scheme in India. Using district-level data on land ownership distribution in 2005 and NREGA implementation during 2006-2010, I find that the concentration of land ownership causes the reduction of public works provision. OLS estimates sug-

Table 1.11: Do districts with more equal land distributions have worse labor markets?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Ag wage	Ag productivity	Backward Index	Literacy Rate	% Marginal Worker	% Ag Labor	% Electr	% Paved Road	% School
Panel A: Full sample									
Gini coef.	-0.199*	0.263	-0.068	0.083*	0.015	0.053	0.198**	0.124**	0.162**
	(0.111)	(0.270)	(0.150)	(0.046)	(0.015)	(0.038)	(0.088)	(0.053)	(0.082)
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1224	1224	1224	1224	1224	1224	1224	1224	1224
R square	0.61	0.27	0.42	0.43	0.30	0.54	0.56	0.81	0.44
Panel B: IV sample									
Gini coef.	0.258	0.138	0.393	0.147*	0.027	0.040	0.049	0.235	0.286**
	(0.200)	(0.516)	(0.366)	(0.077)	(0.037)	(0.083)	(0.147)	(0.146)	(0.140)
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	457	457	457	457	457	457	457	457	457
R square	0.60	0.19	0.40	0.46	0.34	0.57	0.64	0.76	0.47

Note: The table shows comparative economic characteristics in districts with differential land inequality. Each estimate is derived by regressing the economic indicator (column title) on Gini coefficient and state fixed effects. Agricultural Wage (Rupees per day in 1995), Agricultural productivity (Rupees/hectare in 1995) and Backwardness Index all come from 2003 Indian Planning Commission Commission et al. (2003). "% of Electr" means the fractions of villages that have access to electricity in the district. The top panel provides the comparison for the full sample; the bottom panel uses the IV subsample.

* $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

Table 1.12: Robustness of main effects to the control of agricultural wages

	Full sample		IV Subsampe			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS
Gini coef.	-0.64**	-0.65**	-1.71***	-1.52***	-5.12***	-5.22**
	(0.27)	(0.28)	(0.56)	(0.57)	(1.74)	(2.03)
log rural area(Sq. km)	0.29***	0.32***	0.63***	0.64***	0.77***	0.77***
	(0.10)	(0.10)	(0.23)	(0.23)	(0.24)	(0.24)
log Rural population	-0.24***	-0.28***	-0.43**	-0.47***	-0.58***	-0.57***
	(0.08)	(0.09)	(0.17)	(0.17)	(0.18)	(0.18)
Literacy rate	-1.02***	-1.08***	-0.43	-0.39	-0.35	-0.36
	(0.34)	(0.34)	(0.65)	(0.65)	(0.78)	(0.78)
Wet season rainfall deviation	-0.09**	-0.09**	-0.07	-0.07	-0.09	-0.09
	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
% of land covered in fine soil	-0.29	-0.27	-0.94*	-0.76	-2.52***	-2.58***
	(0.21)	(0.21)	(0.52)	(0.51)	(0.82)	(0.99)
% of land covered in medium soil	-0.40**	-0.37**	-0.56*	-0.41	-1.51***	-1.56**
	(0.16)	(0.16)	(0.34)	(0.35)	(0.51)	(0.63)
% of Agricultural labourers	1.73***	1.62***	2.39***	2.31***	3.27***	3.30***
	(0.46)	(0.46)	(0.82)	(0.83)	(1.01)	(1.06)
% of Main workers	-0.23	-0.43	-0.29	-0.25	1.19	1.21
	(0.74)	(0.75)	(1.80)	(1.77)	(2.06)	(2.09)
% of Marginal workers	1.25	1.27	-1.49	-1.61	-1.21	-1.18
	(0.79)	(0.78)	(1.69)	(1.65)	(2.07)	(2.11)
% of SCST population	0.93***	0.87***	0.34	0.34	-0.01	-0.02
	(0.27)	(0.26)	(0.52)	(0.51)	(0.59)	(0.59)
% villages with Safe Drinking water	1.45	1.55	1.04	0.89	-0.29	-0.28
	(1.22)	(1.21)	(3.44)	(3.42)	(3.61)	(3.63)
% villages with Electricity (Power Supply)	0.53*	0.49*	0.55	0.52	0.46	0.47
	(0.28)	(0.28)	(0.42)	(0.41)	(0.58)	(0.59)
% villages with Paved approach road	-0.63**	-0.57**	-0.88*	-0.83*	-0.21	-0.21
	(0.28)	(0.28)	(0.45)	(0.44)	(0.56)	(0.57)
% villages with Primary school	0.08	0.02	1.19***	1.04**	1.52***	1.56***
	(0.33)	(0.33)	(0.43)	(0.47)	(0.45)	(0.49)
% villages with Medical facility	-0.50	-0.46	-0.56	-0.47	-0.56	-0.58
	(0.40)	(0.40)	(0.46)	(0.49)	(0.45)	(0.46)
% villages with Post and telephone facility	-0.36	-0.33	-1.24***	-1.20***	-1.27***	-1.27***
	(0.25)	(0.25)	(0.41)	(0.41)	(0.40)	(0.39)
Phase 2 indicator	-0.18**	-0.16**	-0.34***	-0.30**	-0.37***	-0.38**
	(0.07)	(0.07)	(0.12)	(0.13)	(0.14)	(0.16)
Phase 3 indicator	-0.53***	-0.49***	-0.54***	-0.47***	-0.55***	-0.56***
	(0.08)	(0.09)	(0.14)	(0.16)	(0.16)	(0.18)
Composite Backwardness Index	0.10	0.15*	0.14	0.12	0.42**	0.43**
	(0.09)	(0.09)	(0.17)	(0.16)	(0.18)	(0.20)
Ag Wages (Rs/day, 1996)		-0.29**		-0.35		0.08
		(0.13)		(0.24)		(0.35)
State Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1224	1224	457	457	457	457
Kleibergen-Paap rk Wald F statistics					15.02	11.52

Note: Column 1 restates the OLS result from Table 1.4. Column 3 and 5 restate the main results from Table 1.7.
* p < 0.10, ** p<0.05, *** p<0.01.

gest that a 1 percent difference of land Gini coefficient leads to a 0.6 percent gap in NREGA provision in terms of household participation rates. To address the potential endogeneity issue arising from measurement errors and unobserved omitted variables, I use a historical institution as the instrument variable for land inequality—the land revenue collection system established by British colonial rulers during 1750-1861. First-stage results show that previously landlord-dominated areas in British India has a lower land inequality today, which is because landlord districts, although starting with higher land inequality in British Indian period, enacted a greater number of land reforms after Indian Independence. Under the assumption that the instrument is exogenous, the IV estimates confirm the negative effect of land ownership inequality on public works schemes.

Both OLS and IV results are robust to the use of three alternative measurements of public works provision: per capita labor expenditure, the average days each person in Schedule Caste or Schedule Tribe provided with NREGA employment, and the average number of projects each rural person completes. The results are also robust when I use the shares of land owned by the top 10% largest farmers to measure land inequality, which more directly captures the top distribution and large landlords' political power. To examine the sensitivity of the estimates to the exogeneity restriction of the instrument variable, I construct bounds for the estimated effect by applying Conley et al. (2012). This sensitivity analysis shows the IV results allow for a negative association between the instrument variable and NREGA provision, and a slight positive relation between the IV and NREGA provision. Finally, I exclude the possibility that the higher provision of public jobs in the more equal areas is driven by a higher demand for public jobs, by showing that the local labor market (especially agricultural wages in the private sector) is not worse than that in the more unequal areas.

Investigating the relation between land inequality and public works provision is not only relevant to India, but also has policy implementations for other developing countries, such as South Africa and Kenya, that have the dual need for job creation and investment in public services (such as road maintenance). More broadly, this paper adds to the discussion on how power asymmetries could hinder policies aimed at promoting equity. To improve policy effectiveness, the government needs to take into account asymmetries in bargaining power, which point is highlighted in 2017 World Bank Report (López-Calva et al., 2017). Future research would provide a more complete understanding of the economic consequences of land inequality, and examine how power asymmetry begets economic inequality.

CHAPTER 2

DOES PARTICIPATING IN PUBLIC WORKS INCREASE WAGE BARGAINING POWER IN PRIVATE SECTORS? — EVIDENCE FROM NATIONAL RURAL EMPLOYMENT GUARANTEE SCHEME IN INDIA

Yanan Li (Cornell University)
Yanyan Liu (International Food Policy & Research Institute)

2.1 Abstract

This paper estimates labor market effects of public works programs for participating households in the context of the Indian National Rural Employment Scheme (thereafter, NREGA). Our research question is twofold. First, does working in a public work program increase an individual's own wage bargaining power in private sectors (mostly as agricultural labor)? Second, do husbands' (or wives') participation increase their spouses' wage bargaining power in private sectors? We use a Difference-in-Differences method to estimate the NREGA's effect on participating households' labor market outcomes.

Results show that men would receive a 7% higher wage and work 6 days less in private market if they participate in NREGA program in the main agricultural season; and at the same time, their wives who are not working in public works would tend to reduce labor supply by about 6 agricultural working days and gain agricultural daily wage 6% higher in the private labor market. This result is consistent with a unitary household utility model and wage bargaining story: when husbands participate in the public works program, the benefit obtained

from this program may be transmitted to their wives, hence leading to a higher reservation wage for the latter.

We also find heterogeneous effects by season and by participation intensity. Men's own wage effect and spousal wage effect only exist in the main agricultural season, not in the off season, which means NREGA may bring competition for labor with private sectors in the main agricultural season. Another interesting pattern is that, as husbands work more days and receive more payment from NREGA work, wives' labor supply shows a stronger negative effect. This pattern might indicate an income effect underlying these wage effects.

2.2 Introduction

Public works schemes are an important and widely used anti-poverty policy in developing countries, aside from cash transfer. Existing studies have documented different socioeconomic effects of public works, such as poverty targeting effectiveness, cost-benefit analysis, agricultural productivity, children human capital investment, labor market effects and so on (e.g. Subbarao, 1997; Del Ninno et al., 2009; Zimmermann, 2014; Imbert and Papp, 2015; Muralidharan et al., 2016; Shah and Steinberg, 2015; Azam, 2011; Berg et al., 2014). Among these studies, several have documented a positive wage effect of introducing a public works program (Imbert and Papp, 2015; Azam, 2011; Berg et al., 2014). An explanation for a positive wage effect is that, by providing the rural poor with unskilled employment, especially in the agricultural lean season, public works programs help workers to negotiate for a higher wage and better work environment with rural landlords who are usually oligopolists in the rural labor market (Gaiha, 1996).

Despite the finding of a positive effect on private sector wages brought by public works programs, there is little direct evidence showing the mechanism of how workfare policy affects rural workers' bargaining position. Ideally, we want to answer this question by estimating the equilibrium parameter in a Nash Bargaining game between workers and private sector employers (e.g. Card et al., 2014). However, doing so requires matched employer and employee data, and such data is usually not available in developing countries. In fact, even in developed countries, studies also use indirect approaches to study wage bargaining rather than estimate a Nash bargaining model (e.g. Ballot et al., 2006). Along this line, the current paper provides an implicit test of the bargaining story by estimating Average Treatment Effect on private sector wages for households participating in public works schemes.

We examine whether participating in public work opportunities increases workers' own bargaining power in private sectors and that of their spouses. We hypothesize that, public works programs, serving in the role of unemployment insurance, may pose an upward pressure on private sector wages via higher reservation wages. Therefore, we should observe a positive wage effect in the private agricultural market for program participants. Although we do not directly observe the bargaining parameter, such evidence of a positive wage effect would indicate a potentially higher bargaining power for program participants. If we further assume that husbands and wives share resources within the household, then public employment would also increase spouses' wage bargaining power in the private labor market.

The empirical framework employs a difference-in-differences model in the context of India's Mahatma Gandhi National Rural Employment Guarantee Scheme (hereafter, NREGA program). The identification assumption is that the distribution of NREGA job opportunities is exogenous to households. In other words, without NREGA job individual wage growths in participating and non-participating households would have identical trends. This study takes advantage of a unique data set, a combination of a household survey panel in 2005-06 and 2007-08 and Indian administrative data that have NREGA participation records. Compared to most existing studies that use repeated cross-sectional NSSO employment data, the advantage of a household survey panel is that it allows to control for individual specific time-invariant unobservables. Moreover, with seasonal variations of labor market participation, we can get estimates for pre-treatment trends.

NREGA is the world's largest public program so far according to World Bank report in 2015. It provides at least 100 days of guaranteed wage employment in a financial year to each household whose adult members volunteer to do unskilled manual work at the minimum wage level. It focuses on unskilled work such as water conservation, drought proofing, irrigation works and land development. Starting from February 2006, the program had gradually expanded throughout India by mid-2008. Like other public works program, NREGA is designed to help the poor stabilize income and smooth consumption in the agricultural off-peak season. About 10% of the rural labor force works in this program sometime during the year.

Due to the self-targeting goal of NREGA, program participation is a result of self-selection by design. This selection issue undoubtedly poses challenges to identifying wage effects. For instance, if poorer people are more likely to work in the program and if they have different wage and employment paths from richer people, then the common trend assumption underlying dif-in-dif model may not hold. We utilize the empirical approach used in the well known analysis of job displacement by Jacobson et al. (1993). This methodology allows to simultaneously estimate all pre-treatment trends of outcomes in addition to main treatment effect in current periods. If participants and nonparticipants present similar wage growth paths prior to the introduction of NREGA, then our estimation is less likely to be driven by self-selection.

In our main findings, we find that if husbands participate in NREGA in agricultural main season, they would gain a 6 percent wage increase in the private agricultural labor market. At the same time, their wives who are not working in public works tend to reduce labor supply by about 6-10 agricultural working days, and gain 7 percent higher daily wage in private labor market. This result is consistent with a unitary household utility model and wage bargaining story. Intuitively, when husbands participate in public works program, the benefit obtained from this program may transmit to their wives as well, hence leading to a higher reservation wage for the latter.

Another two interesting findings include heterogeneous effect by season and by participation intensity. Specifically, men's own wage effect and spousal wage effect only exist in agricultural main season, not in off season. The rational is that, in Karif/Rabi season there is already a relatively large labor demand in private sectors, thus the introduction of NREGA program brings competition for labor against private sector. In contract, in Summer season, labor demand is low in private labor market, so NREGA does not result in competition with the private market. The other interesting pattern is as husbands work more days and receive more payment from NREGA work, wives' labor supply show a stronger negative effect. This pattern may indicate income effect underlying these wage effects. On the other hand, women's own response is in similar magnitude but only appears in agricultural off season. And husbands do not respond to wives' participation.

This paper relates to the literature of labor market impacts of workfare schemes in low-income countries (see Devereux and Solomon, 2006). Several studies have documented a positive earnings (or wage) effect of NREGA in agricultural labor market (e.g. Basu et al., 2009; Berg et al., 2014; Imbert and Papp, 2015; Azam, 2011). They find government hiring via public works programs may crowd out private sector work and therefore leads to a rise in equilibrium private sector wages. However, some other studies find zero or marginal earnings effect (e.g. Zimmermann, 2012). To differentiate our work from existing literature, we need to distinguish two concepts — Average Treatment Effect (ATE) and Average Treatment Effect on the Treated (ATT). The former averages treatment effect for both compliers (or participants) and noncompliers (or non-participants in program available areas), compared to NREGA-non-available areas. In the context of public works program (which usually has spillover effect), ATE tells two things. First, it provides a lower bound of wage effects for the real participants (or ATT). Second, similar to general equilibrium wage effects, ATE proves estimates for wage effects regardless of participation status. In other words, even for those who do not participate in the program, the presence of public program still has an option value of increasing reservation wages. While most existing studies have estimated Average Treatment Effect on private wages of public works program (e.g. Imbert and Papp, 2015; Zimmermann, 2012), this paper estimates Average Treatment Effect on actual participants to shed light upon wage bargaining effect.

Second, our paper is analogous to the literature of unemployment insurance in developed countries. It is a long debate whether unemployment insurance reduces labor supply and increases reservation wages. Using censored regression model and Heckman two-stage estimation method, previous studies find that reservation wages of the unemployed decline 0.6 percent over time, and drop by 15 percent when benefits are exhausted (e.g. Kiefer and Neumann, 1979; Fishe, 1982). Consistent with the literature, our paper finds that participation in NREGA increases both women's and men's agricultural wages in the private labor market and reduces wives' labor supply.

This paper also relates to a few other literature. The wage bargaining effect adds a new dimension to the documented welfare effects of NREGA, including poverty reduction Ravi and Engler (2015), welfare redistribution from rural labor employers to workers Imbert and Papp (2015) and general welfare (e.g. Basu and Sen, 2015). By studying how spouses respond to the partners' participation in public works program, our paper directly speaks to the literature of Added Worker Effect in developed countries.

The rest of paper is organized as below. Section 3 provides background information of NREGA program implementation. Section 4 builds a theoretical framework for this analysis. Section 5, data. Section 6, empirical model and identification. Section 7, results. Section 8 concludes and discusses future work.

2.3 Program Background

Here are some relevant facts about this program. NREGA is a three-phase roll-out program, with 199 districts in Phase 1 (Feb 2006), 128 districts in Phase 2 (April 2007) and the remaining 261 districts in Phase 3 (April 2008).

This program issues a unique job card two weeks after they apply for NREGA works and get approved. Job cards are then used to keep track of days worked and payments received by each participant. A job card identification number also contains the information where the household resides in, such as state, district and village. Job card information is publicly available in NREGA official website to protect labors against corruption and fraud.

Several households may apply for a project and then work on it together, such as irrigation, road pavement etc. Within a household, more than one member can work in the project at the same time.

2.3.1 Wage and Rationing of NREGA work

The average daily wage on NREGA work is 81 Rupees, as opposed to about 55 Rupees/day for women and 86 Rupees for men working as agricultural casual labor (mostly casual labor hired by landlords).[1] Thus, NREGA work is usually seen more attractive than working as agricultural casual labor in private sector, especially for women. This is consistent with the initial aim of this program – to empower women by proving them employment opportunities.

Although the program asserts providing 100 days working opportunity for

[1]Authors' calculation based on our sample

each household per year, there is actually an unmet demand of work. The average working days is roughly 35 days for all members of the household during that year.[2] The rationing of demand for NREGA work is a reason that across Indian states the number of NREGA days provided is only weakly correlated with poverty (Dutta et al., 2012).

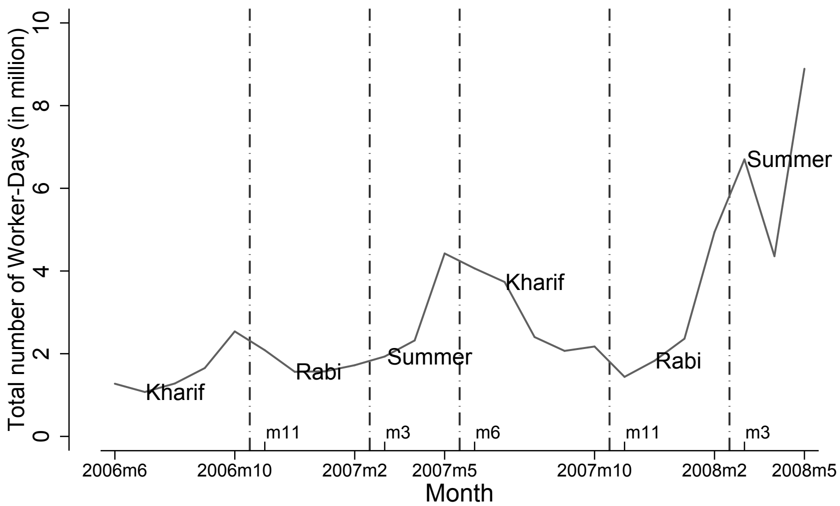
In terms of workers' time allocation, most of those (above 50% based on our survey data) who participate in NREGA work as agricultural or non-agricultural casual labor in private sector, with only a small fraction of them work in salary jobs.

2.3.2 Seasonality of NREGA works

There are three main agricultural seasons in India, i.e. Karif (June-Oct), Rabi (Nov to Feb) and Summer season (March to May). Karif season is concurrent with monsoon season, hence agricultural busy season, and has a relatively large casual labor demand by landlords. The competition of private sector and public sector for rural labor makes it possible for a positive wage effect of this program. Rabi season is winter season with less labor demand in private agricultural sector. Summer season is very dry and hence agriculture lean season with little labor demand by landlords. The introduction of NREGA program helps to stabilize labor demand in lean seasons.

Figure 2.1 presents the seasonality of NREGA works in our survey districts in Andhra Pradesh state. The number of worker-days varies by season and month. To avoid competition with private sector labor demands, NREGA pro-

[2] Authors' calculation based on our sample



Note: Agricultural seasons are indicated between vertical lines.
Kharif season=June-Oct; Rabi season=Nov-Feb; Summer season= March-May
The author calculates total number of worker-days by month since the introduction of NREGS.
Data is downloaded from muster records in NREGS website.
The sample is restricted to 5 districts that our survey covers in AP rather than all all districts.

Figure 2.1: Seasonality of NREGA works, 2006.6-2008.5

gram provides more works in off-agricultural season and less in agricultural busy season. This pattern in our data is consistent with existing studies (e.g. Maiorano, 2014; Imbert and Papp, 2015).

2.4 Modeling and Hypothesis

We use the framework of McCall (1970)[3] to show why the introduction of public works program could increase participants' reservation wages.

Use w_r to represent reservation wage, and w actual job offer, b is the income one can get if not working in private sector

$$w_r - b = \frac{\beta}{1-\beta} \int_0^{\infty} (w - w_r) dF(w) \tag{2.1}$$

[3]These two notes are helpful. http://lhendricks.org/econ720/search/McCall_SL.pdf and http://people.hss.caltech.edu/~kcb/Notes/JobSearch.pdf

To rearrange it,

$$b = w_r - \frac{\beta}{1-\beta}\int_0^\infty (w - w_r)dF(w) \tag{2.2}$$

In the context of a public program that guarantees some employment or cash-on-hand, the utility (in terms of income) that one can get from opting out of private sector increases if one participates in NREGA program. Note also that the RHS of equation 2.2 is a monotonically increasing function of w_r. As a result, Participating in NREGA increases reservation wage.

While looking at spousal response to partners' participation in public works program, we need to assume a unitary household model and intra-household sharing mechanism — the benefit from NREGA program may transmit from participants to non-participant members in the same household. Compared to individuals from non-participating households, these non-participants from treated households have better fallback options, hence more likely to have a higher bargaining power in negotiating wages with landlords in private labor markets.

2.5 Data

Our sample includes 471 villages in 5 districts in Andhra Pradesh, i.e. Visakhapatnam, Nellore, Kadapa, Warangal and Nalgonda. Our data comes from three sources. First, Rural Poverty Reduction Project survey data in 2004, 2006 and 2008 agricultural year; second, NREGA administrative data from the official website; third, Indian population census data.

The survey data contain NREGA job card Identification Number and detailed information of household members' labor market participation (other than in NREGA programs), such as demographic backgrounds and salary or wage in each work by season. 2004 survey was the first wave survey data, mostly conducted during March-August 2004. The interview asks the subject to recall information during June 2003-May 2004. Then, 2006 survey was conducted intensively during August and October 2006; subjects were asked to recall information during June 2005-May 2006. Similarly, 2008 survey was conducted during September-December 2008, and subjects recalled information between June 2007-May 2008. Our survey data almost two waves of survey data prior to the introduction of the program, and one wave after.

The administrative data (muster rolls) is downloaded from nregs official website. It contains information on job card identification number, NREGA participation for each participant, such as the start and end date of working at a specific project in NREGA program, and total payment during each recorded working period. Because our survey data is at person-season level, we need to aggregate NREGA participation information into season level as well.

Population census data contain village information such as rainfall and other village characteristics.

Since both survey and administrative data has job card information and individual names, we use these to merge survey households and NREGA-participating households from administrative data. The final data is in the form of household-member-season. For each member in the household, we have labor market participation information in each season.

2.5.1 Program roll out and take-up

Table 2.1 documents how NREGA program rolled out in our sampling villages and the variation of program take-up. Our survey divides the year into three agricultural seasons based on rainfall amount, i.e. Karif season =June-October, Rabi season = November-Feb, Summer season=March-May.

The start of NREGA program in a village is defined by the first day that any household starts to work in this public program. In other words, suppose NREGA program is already available in a village and households can apply for it, but none of them really do, hence no NREGA work is going on in the village, then this village is still viewed as a non-NREGA village. In this way, we find the rolling out process of this program at village level. Our sample contains 471 villages in 5 districts. Table 2.1 shows at the end of the survey window, only 45 villages still didn't have access to NREGA.

Table 2.1: Program phased roll-out at village and individual level

Survey year	season	Villages			Individuals		
		Starting NREGA	With NREGA	Without NREGA	# of non-part.	# of participants	participation rate
2006	Kharif	0	0	471	8509		
2006	Rabi	2	2	469	8494		0
2006	Summer	219	221	250	8342	68	1.90%
2007		75	296	175			
2008	Kharif	42	338	133	8156	779	12.50%
2008	Rabi	11	349	122	8254	664	9.81%
2008	Summer	77	426	45	7663	1,165	15.97%
	post survey	45					
	total # of villages	471					

Table 2.1 also suggests NREGA takes a long time to take off, when we compare village roll out and households take up rate. Although half of the villages already had access to NREGA in May 2006 (phase 1), only 2% individuals actually worked in it. Phase 2 districts started in April 2007. Our data does not cover this period. Starting in June 2007, take up rate increased to around 12.5% in our sampling villages.

We exploit the fact that this program was taken up gradually at individually level, treating three seasons in 2006 survey year as pre-treatment periods, and the corresponding seasons in 2008 as post periods.

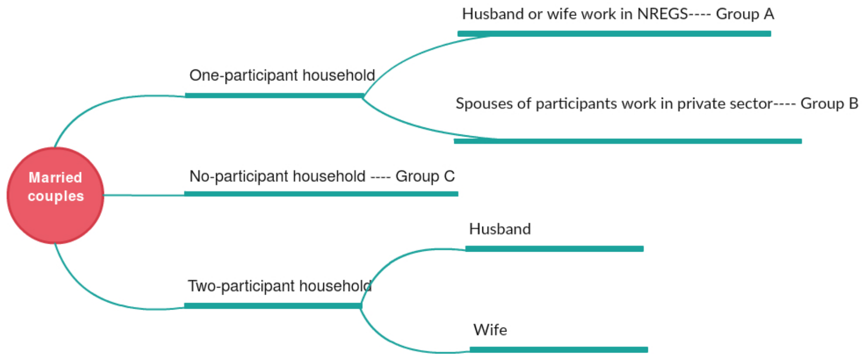


Figure 2.2: Grouping of the sample

2.5.2 Descriptives

Table 2.2 presents a comparison for three groups of individuals based on their own participation status in NREGA program and their spouses' participation status. Karif season and Rabi season are aggregated as agricultural main season. First I divide all households in the sample into three types depending on couples' participation status in NREGA program, see Figure 2.2. Type 1 households, or "one-participant households", have either wife or husband participate in NREGA program; Type 2 households, or "no-participant households", have neither of the spouse participate in NREGA; Type 3 households, or "two-participant households", have both of the spouse participate in it. Second, I further divide individual workers from "partially-participating households" into two groups – participants and non-participants, as shown in block 1 and block 2 in Table 2.2.

We will estimate the spillover effect of participating NREGA by comparing non-participating spouse from partially-participating households and workers from non-participating households. The last two blocks in Table 2.2 presents the comparison for these two groups. Panel 2 presents average number of days

Table 2.2: Descriptive statistics – three group comparison

	Part. from partial-part. Households				Non-Part. from partial-part. households				Individuals from non-part households				Part. from all-part. Households			
	Main season		Summer season		Main season		Summer season		Main season		Summer season		Main season		Summer season	
	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband	Wife	Husband
Total # of observations	356	200	201	116	200	356	116	201	3457	3457	1955	1955	327	326	291	290
NREGS work and payment																
working days/season	13.5	10.3	14.0	9.5												
payment (Rupee/season)	981.4	755.3	977.7	644.0												
unit payment (Rupee/day)	71.5	71.1	67.7	66.4												
Working days by job type																
Ag casual labor (days)	65.2	53.7	30.9	31.9	57.7	57.3	27.5	32.3	58.4	55.3	32.9	34.2	61.3	54.1	28.2	27.0
# of workers	280	105	139	61	128	163	65	68	2062	1424	902	695	273	217	204	147
non-Ag casual labor (days)	16.2	20.8	22.1	17.6	30.9	45.0	22.4	37.5	27.6	44.4	23.6	33.1	16.4	19.5	21.7	22.3
# of workers	356	200	201	116	49	105	51	89	596	782	580	638	327	326	291	290
day_salary (days)		138.0		85.2	112.0	125.6	90.5	87.7	114.5	128.2	84.9	88.2	91.5	126.1	66.4	76.4
# of workers	0	11	0	5	3	43	2	22	74	312	39	180	2	16	5	17
day_selfemp (days)	42.5	85.0	24.2	53.5	88.8	81.9	60.0	59.9	82.7	85.6	60.6	63.0	67.5	75.4	40.9	50.3
# of workers	8	15	11	10	12	35	7	28	294	504	161	268	15	30	9	20
Wage by job type (Rupee/day)																
Ag casual labor	50.0	79.3	48.6	67.4	48.7	78.6	43.9	74.1	48.0	76.7	46.3	74.7	49.3	76.5	46.6	69.5
non-Ag casual labor	71.0	73.0	67.0	68.9	66.9	86.8	62.5	85.6	68.1	89.7	64.8	79.3	69.4	75.3	67.3	71.3
salary wage		2034.8		1900.0	1466.7	3285.8	2850.0	3371.3	1937.1	3149.9	1635.4	3507.9	3250.0	4675.0	1650.0	2281.3
self emp wage	50.0	113.3	53.2	98.9	217.5	134.0	71.4	153.0	123.4	214.2	127.8	241.5	62.1	175.5	275.0	214.2

The sample is divided into three types of households, based on individual participation status in NREGA program and their spouses' participation status. Type 1 households, or "one-participant households", have either wife or husband participate in NREGA program; Type 2 households, or "no-participant households", have neither of the spouse participate in NREGA; Type 3 households, or "two-participant households", have both spouses participate in it. Then type 1 individual workers from "partially-participating households" into two groups – participants and non-participants, as shown in block 1 and block 2 an individual works if he/she does that type of work. Panel 3 is informative in terms of potential wage effect. For instance, in agricultural season, a female non-participant from treated households on average earns 48.7 Rupees/day, as opposed to 48 Rupees/day for a female worker from control households. In addition, the former works on average 57.7 days as agricultural wage labor, as opposed to 58.4 days in the latter group. The fact that non-participating spouses from treated households receive a higher wage and work fewer days than individuals from control households is consistent with our empirical results.

The first block about participants from partially-participating households in Table 2.2 answers the following questions. First, it shows how many days and how much do they earn in each season. Females participate in NREGA for more days than males, have equal daily payment, and on average earning more than males. Such results are consistent with the initial goal of empowering women. Second, it answers what else these participants do other than NREGA works. Quite surprisingly, Panel 2 shows that they work for no fewer days than people not participating in NREGA, consistent with an earlier finding of unmet demand of nregs works. Panel 3 shows their wages in private sector are a bit higher than non-participants.

According to National Sample Surveys (NSS), in 2006-07, the average monthly per capita expenditure (MPCE) for rural households is 695 Rupees or about $14. About 52 percent of this MPCE was spent on food[4]. NSSO survey on situation assessment of farmers (2003) estimate that a farmer household, on the average, has a total monthly income of 2115 Rupees from all sources (Bahala, 2008).

[4]http://www.prb.org/Publications/Articles/2008/howindianslive.aspx

2.6 Empirical Model and Identification

In a village with NREGA program, some households apply for and finally get work opportunities from this program, whereas other households may either not apply or finally do not pass final review process. We define the first type of households "participating households" where either husband or wife (but not both) participates in NREGA program, and the second type "non-participating households" if neither husband nor wife participates in the program. We have dropped households where both husband and wife work in NREGA program, because in those families it is unclear whether individual i's wage change is a reaction to its own participation to the program or a reaction to its spouse's participation.

NREGA gradually rolls out to 426 out of 471 villages in our sample areas during 2006-2008. To estimate ATT on wages, theoretically there are two different comparisons we could make. One is comparing participating households in NREGA-available villages to households in NREGA-non-available villages, and the other is comparing participating households to non-participating households in NREGA-available villages. While the former comparison is the conventional way of estimating ATT, instrumental variable methodology needs the assumption that the rolling out process needs to be random across villages. Alternatively, the variation of participation status in the second comparison purely comes from individual self selection, which poses a threat to identify the wage effect. Considering the fact that random assignment of NREGA at village level seems too strong an assumption, we use the second comparison and try to identify ATT by mitigating the concerns due to self-selection. Because even non-participants are also faced with the "option value" of the availability of the program, this comparison will yield an underestimated wage effect.

The identification strategy for ATT is based on the assumption that the distribution of NREGA job opportunities is exogenous to households, so that without NREGA job, individual wage growths in Treatment and Control households would have identical trends. However, if some households (e.g. elite class) have manipulation power on the distribution of job opportunities, then this assumption will be violated. For instance, if households with high-skill nonparticipants are more likely to obtain NREGA work opportunities, then the effect of receiving public works on non-participants' private sector wages will be confounded by non-participants' skill/ability.

Fortunately, with seasonal data in our sample, we can use the model in the well cited job displacement study by Jacobson et al. (1993) to identify self response and spousal response of labor market outcomes to participating in NREGA for the participating households. It's essentially a dif-in-dif framework but allows to simultaneously estimate all pre-treatment trends of outcomes in addition to main treatment effect in current periods. If participants and nonparticipants present similar wage growth path prior to the introduction of NREGA, then our estimation is less likely to be driven by self-selection.

Empirical specifications for self-response analysis are as follows. In estimating individual response to its own participation in the program, we define the treatment indicator D_{it} as follows: $D_{it} = 1$ if individual i works in NREGA program at time t, and otherwise $D_{it} = 0$. I estimate the same model separately for wives and husbands. In the parallel analysis of spousal response to the partner's participation, treatment indicator is defined in the same way, but left hand side variables in the regression model are the spouse's outcomes rather than individual i's own outcomes.

We have two years of data, 2006 and 2008, each with three seasons. In 2006, no one gets treated. In 2008, participants move in and out of NREGA program during the three seasons. Therefore, we could compare labor market outcomes for participants and non-participants in each season in 2008, using three seasons in 2006 as three pre-treatment trends. For instance, to estimate the treatment effect for Karif season in 2008 ($p =$ Karif season, 2008), I extract the sample that appear in Karif season of 2008 and also in the three seasons of 2006. Then I estimate the following model:

$$y_{it}^{p} = \left(\sum_{k=-2,-1,0} \gamma_k^p D_{it}^k \right) + \alpha_i^p + \lambda_t^p + \beta^p X_{it} + \varepsilon_{it}^p \tag{2.3}$$

where t indicates time, one of the seasons in 2006 and the treatment time in 2008. α_i^p captures individual fixed effect; λ_t^s captures seasonal time trends. X_{it} includes age squares and reading ability for individual i and its spouse, caste and dependency ratio interacted with time. In the self-response analysis, y_{it}^p is individual $i's$ own wage and workdays in the private agricultural labor market; $D_{it}^k = 1$ for NREGA participant $i's$ k^{th} season from the current treatment time p (and in this example $p =$ Karif season, 2008). Note that the three seasons in 2006 serve as pre-treatment periods. $k = -2$ stands for Rabi season in 2006, the second to last season before the 2008 Karif season. $k = -1$ stands for Summer season in 2006, the last season before the current treatment season. The first season in 2006, Karif season, is used as the baseline and omitted in the regression. $k = 0$ stands for the current season, and in this example Karif season in 2008. Therefore, γ_0^p gives the estimate for the effect of participating in NREGA; γ_{-1}^p and γ_{-2}^p give the estimates for pre-treatment effects and provide the test for common pre-trend assumption. If estimates for γ_{-2} and γ_{-1} are close to zero, then it supports the common pre-trend assumption.

Similarly, we could estimate the treatment effect for participants who work in NREGA in Rabi or Summer season in 2008 using the same model. However, the drawback of doing these three estimations separately by season is that it is hard to get statistical significance, because of the low take-up rate in the program. Therefore, we append these three samples together, and estimate Equation 2.4. Because both Karif and Rabi seasons are agricultural main seasons with substantial agricultural labor demand, we further assume the treatment effects for these two seasons are homogeneous. Eventually, the seasonal heterogeneous effect is split into agricultural main and off season. All results given in the paper are based on this combined model[5]:

$$y_{it} = \left(\sum_{k=-2,-1} \gamma_k D_{it}^k \right) + \gamma_m D_{it}^m + \gamma_l D_{it}^l + \alpha_i + \lambda_t + \beta X_{it} + \varepsilon_{it} \tag{2.4}$$

where γ_m and γ_l provide the estimated effects for agricultural main season and lean season, respectively.

2.7 Results

For all the results reported below, it seems that statistical significance is not quite high. This is because of the small number of treated individuals. Having dropped families where both spouses participate in NREGA, the final sample contains on average 90 men or 150 women participants in each season. The following sections present individual own response and spousal response, and then a pattern of these effects.

[5]We find similar results when estimating treatment effects for each season separately following Equation 2.3.

2.7.1 When husbands work in NREGA

Table 2.3 presents men's own response to participating in NREGA, and Table 2.4 presents wives' labor market response to husbands' participation. Agricultural labor only happens in private market, whereas casual labor includes both agricultural labor and non-ag labor (note that NREGA work mostly is recorded as non-ag casual labor). In each work type, the outcome variables are wage, intensive margin of labor supply (i.e. how many days worked) and extensive margin of labor supply (i.e. whether or not working) Essentially, we would like to see when a husband participates in NREGA, whether he works more or less in private labor market and how his private market wage changes. We also want to see how his wife, who does not work in this NREGA but only participates in private market, changes her labor market performance.

Table 2.3: Men's Self Response, using NREGA payment 300 Rupees as cut off

	Casual labor			Ag casual labor		
	(1)	(2)	(3)	(4)	(5)	(6)
	Wage	Days	Work Y/N	Wage	Days	Work Y/N
Treatment * Rabi season, 2006	-0.01	-0.02	-0.01	-0.00	-1.29	-0.00
	(0.01)	(1.79)	(0.01)	(0.01)	(1.66)	(0.01)
Treatment * Summer season, 2006	0.01	2.97	-0.00	0.01	0.01	-0.01
	(0.01)	(2.29)	(0.02)	(0.01)	(2.28)	(0.02)
Treatment * Main season	0.02	3.63	0.05	0.07*	-5.20	0.02
	(0.04)	(4.83)	(0.04)	(0.04)	(4.66)	(0.05)
Treatment * Off season	0.01	4.17	0.09	0.00	2.83	0.01
	(0.04)	(4.37)	(0.06)	(0.05)	(4.70)	(0.06)
Observations	14790	14790	29532	12275	12275	29532
R^2	0.816	0.715	0.725	0.810	0.706	0.697

All models include a full set of individual and season fixed effect, and observable covariates. Standard errors are clustered at household level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 2.4: Spousal response, wife to husband, NREGA payment 300 Rupees above

	Casual labor			Ag casual labor		
	(1)	(2)	(3)	(4)	(5)	(6)
	Wage	Days	Work Y/N	Wage	Days	Work Y/N
Treatment * Rabi season, 2006	-0.01	-1.48	-0.02	-0.01	-1.54	-0.01
	(0.01)	(1.75)	(0.01)	(0.01)	(1.77)	(0.01)
Treatment * Summer season, 2006	0.01	3.57	-0.02	0.02	3.20	-0.01
	(0.01)	(2.72)	(0.03)	(0.01)	(2.83)	(0.03)
Treatment * Main season	0.05	-5.97	-0.07*	0.06*	-6.86	-0.10**
	(0.04)	(4.97)	(0.04)	(0.03)	(4.69)	(0.05)
Treatment * Off season	0.02	4.68	0.02	0.04	0.56	0.02
	(0.04)	(3.99)	(0.05)	(0.04)	(4.28)	(0.06)
Observations	16731	16731	29532	15911	15911	29532
R^2	0.780	0.664	0.713	0.791	0.676	0.694

Notes: Casual labor includes both ag and nonagricultural casual labor who earns daily wage. Column 1, 2 and 4, 5 restrict the sample to individuals who work a positive number of days as agricultural wage labor, whereas column 3 and 6 also include individuals who don't work as agricultural wage labor. All models include a full set of individual and season fixed effect, and observable covariates. Standard errors are clustered at household level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

The first thing we observe from Table 2.3 and Table 2.4 is pre-treatment effect. Table 2.3 shows wage paths of participant husbands do not differ from that of non-participant husbands in terms of statistical significance and economic magnitude. Similarly, Table 2.3 tells wage paths of wives whose husbands participate in NREGA are not statistically different from wage paths of wives whose husbands do not participate in NREGA.

Second, look at wage effect and employment effect. Table 2.3 shows that participant men gain 7% wage increase in agricultural labor market, in agricultural main season. This effect probably indicates that the introduction of NREGA has led to competition for labor between private sector and public works.

Third, corresponding to a positive wage effect for participant husbands, Table 2.4 shows, if husbands work in NREGA, their wives tend to gain 6 percent increase and work 6 days less in agricultural labor market, compared to women whose husbands do not work in NREGA. This is consistent with our story of wage bargaining and unitary household model.

2.7.2 When wives work in NREGA

Different from men's effect, women's effect concentrates in agricultural off season. Table 2.5 shows women's participation as casual wage labor in private sector has increased in lean season. This result indicates that NREGA has helped to generate more employment for women, especially in agricultural off season. At the same time, results also suggest a crowding out effect for female workers in off season. Participant women work for 5 fewer days and earn 8 percent higher wage as agricultural daily worker, compared to non-participant women.

The finding that treatment effects only appears in off season is also consistent with that in Imbert and Papp (2015), although they do not distinguish men and women. As they argue, NREGA work is mainly going on in off season. This reason may also apply in our data, as Figure 2.1 shows more work is going on in Summer season. More interestingly, seasonality is not simply due to variation in demand, but might also reflect elite pressure. For instance, as Anderson et al.

(2015) argue, landlords might seek to control governance in order to suppress wage pressure.

Table 2.5: Women's Self Response, using NREGA payment 300 Rupees as cut off

	Casual labor			Ag casual labor		
	(1)	(2)	(3)	(4)	(5)	(6)
	Wage	Days	Work Y/N	Wage	Days	Work Y/N
Treatment * Rabi season, 2006	0.00	-0.86	0.01	0.01	-0.97	0.01
	(0.01)	(1.26)	(0.01)	(0.01)	(1.19)	(0.01)
Treatment * Summer season, 2006	0.01	-5.97***	-0.03	0.00	-6.67***	-0.04*
	(0.01)	(2.14)	(0.02)	(0.01)	(2.17)	(0.02)
Treatment * Main season	0.04*	3.50	-0.04	0.03	2.24	-0.05
	(0.02)	(2.94)	(0.03)	(0.02)	(2.84)	(0.03)
Treatment * Off season	0.08***	4.94*	0.04	0.08***	-5.45*	0.01
	(0.02)	(2.91)	(0.03)	(0.03)	(3.16)	(0.04)
Observations	16731	16731	29532	15911	15911	29532
R^2	0.780	0.665	0.713	0.792	0.677	0.694

Note: Casual labor includes both ag and nonagricultural casual labor who earns daily wage. Column 1, 2 and 4, 5 restrict the sample to individuals who work a positive number of days as agricultural wage labor, whereas column 3 and 6 also include individuals who don't work as agricultural wage labor.
All models include a full set of individual and season fixed effect, and observable covariates. Standard errors are clustered at household level. * p<0.10, ** p<0.05, *** p<0.01.

Table 2.6 estimate husbands' response to wives' participation in NREGA. Unlike wives' response to husbands, when wives work in NREGA, husbands do not show any statistically significant reaction in either agricultural labor market or casual labor market as a whole. Only in column (5), the effect on intensive working days is close to being statistically significant with t-statistics being 1.4. This means husbands whose wives work in NREGA tend to work for 6 days less than husbands whose wives don't work in NREGA.

Table 2.6: spousal response, Husband to wife, NREGA payment 300 Rupees above

	Casual labor			Ag casual labor		
	(1)	(2)	(3)	(4)	(5)	(6)
	Wage	Days	Work Y/N	Wage	Days	Work Y/N
Treatment * Rabi season, 2006	0.00	-0.37	0.01	-0.01	-1.42	0.00
	(0.01)	(1.41)	(0.01)	(0.01)	(1.47)	(0.01)
Treatment * Summer season, 2006	0.01	-2.45	0.01	0.00	-3.75*	0.02
	(0.01)	(1.98)	(0.02)	(0.01)	(2.14)	(0.02)
Treatment * Main season	-0.01	-4.54	0.00	0.02	-3.19	0.02
	(0.04)	(3.61)	(0.03)	(0.03)	(4.23)	(0.04)
Treatment * Off season	-0.03	-0.92	-0.01	-0.03	-6.01	-0.03
	(0.04)	(3.44)	(0.04)	(0.04)	(4.16)	(0.04)
Observations	14790	14790	29532	12275	12275	29532
R^2	0.816	0.715	0.725	0.809	0.706	0.697

All models include a full set of individual and season fixed effect, and observable covariates. Standard errors are clustered at household level. * p<0.10, ** p<0.05, *** p<0.01.

2.7.3 Pattern of treatment effects

We test if our earlier estimates rely on the definition of treated households. In the context of wage bargaining story, a tiny amount of monetary benefit from the program may not be helpful enough to raise reservation wage. In the main results given above, as long as husband/wife participates in the program and receives more than 300 Rupees, then their households are counted as treated households. In robustness checks, I redefine treated households as, having the spouse work in the program and receive money greater than a certain amount of Rupees. I tried several thresholds, i.e. 100, 200, ..., 800.

An interesting pattern is, as husbands or wives work more days and receive more payment from NREGA work, the according effects are getting stronger. This probably indicates the role of income effect underlying wage and employment response.

2.8 Conclusion and Discussion

This paper uses a difference-in-differences method to estimate labor market effects of public works schemes for the participating households, in the context of NREGA. We find a positive wage effect (and a negative labor supply effect) for the participants of public work schemes and for their spouses. The results are consistent with a unitary household utility model and wage bargaining story.

First, when either husbands or wives work in NREGA, their own wages at agricultural labor market tend to increase by 7 percent and work for 6 days less, although we are not clear why wives' own response mainly occurs in agricultural off season, while men's wage effect is in main season. Second, combining results of self response and spousal response, we find an interesting phenomenon — own and spousal response go side by side. In main season, when participating men gain 7% positive wage effect, hence a possible income effect, their wives who are not working in public works programs reduce labor supply by about 6-10 agricultural working days, and gain 6% higher agricultural daily wage. Similarly, in the lean season, when participating women gain 8% wage increase, their husbands usually work 6 days less than non-participant women's husbands, although not statistically significant. This result is consistent with a unitary household utility model and wage bargaining story. Intuitively, when husbands (wives) participate in public works program, the benefit obtained from this program may transmit to their wives (husbands) as well, hence leading to a higher reservation wage for the latter.

The findings are further supported by the heterogeneous effects by season and by participation intensity. Men's own wage effect and spousal wage effect only exist in agricultural main season, while women's in agricultural off season. This suggests NREGA works may bring competition for labor in agricultural main season. Another pattern is as husbands work more days and receive more payment from NREGA work, wives' labor supply show a stronger negative effect. This pattern may indicate an income sharing mechanism underlying these wage effects.

The identification of our estimates relies on the assumption that, conditional on observables included in our model, the distribution of NREGA job opportunities is exogenous to households. In other words, without NREGA employment, individual wage growths in treatment and control households would have identical trends. By using the multiple-period dif-in-dif methodology used in JLS's job displacement analysis, we try to show that the estimated wage effects are not driven by unobserved pre-treatment trends because the pre-treatment effects are all close to zero.

This identification strategy is, however, prone to the following two main critiques. First, which households are treated? If households self-select into the programs, for instance, if poorer people are more likely to work in the program and if they have different wage and employment paths from richer people, then the common trend assumption underlying dif-in-dif model will be violated. Second, within each treated household, how do the husband and the wife decide which person goes to work in NREGA? If they make joint decisions, e.g. husbands work in NREGA just because their wives are competitive in private labor market, then the positive correlation between husbands' participation in NREGA and wives' private sector wages will be spurious. In addition, the estimation only exploits the sample of "partially-participating households" and "not-participating households", and drops households that have both spouses participate in NREGA.

The first concern could be partly addressed by several facts on job allocations of NREGA. First, during the study period, low take-up rate (around 10%) eases the concern on households' self-selection into the program. Only a small fraction of lucky households participates in NREGA in the first two years, because workers are unaware of their entitlement to employment. This situation improves only when local volunteer organizations help them to learn to apply for a job card, demand work and open a bank account, tracking the payment of their wages and filing complaints. For instance, in Jharkhand state, local volunteers operate the program NREGA Sahayata Kendras to help workers secure work entitlements. Even after rural workers are fully aware of this job opportunity, there is still a considerable un-met demand in all states due to supply constraints (e.g. Dutta et al., 2012; Sukhtankar, 2016). Second, the accuracy of targeting is in general insufficient, as large numbers of needy households are in the queue for job cards (Jha et al., 2008). This means that not only rural poor but rural non-poor also participate in NREGA, although participation rates among the poor are indeed higher (Dutta et al., 2012). Inefficient targeting and supply constraint to some extent mitigate the concern on households' self-selection into the program by poverty status or individual capability of finding a job in private sector.

Furthermore, I will conduct two sets of empirical analysis to address these two concerns. The first analysis provides an evaluation of the magnitude and direction of household self-selection. As indicated in Figure 2.2, there are four types of households. Type 1 households have only husbands participate in NREGA program; Type 2 households have only wives participate in NREGA program; Type 3 households, "no-participant households", have neither of the spouses participate in NREGA; Type 4 households, "two-participant households", have both of the spouses participate in it. For each type of households, I create a binary indicator D_h^m, which equals to 1 if the household h belong to Type $m(m = 1, 2, 3, 4)$ and 0 otherwise. Then I examine the likelihood of households self selecting into the program by estimating the following equation:

$$D_{ht}^m = \beta^m X_{ht} + \lambda_t^m + \varepsilon_{ht}^m \tag{2.5}$$

where t indicates one of the six seasons in 2006 and 2008; X_{ht} is a vector of household characteristics, including caste, household poverty type, religion, age of the household head, reading and writing ability of the household head, household dependency ratio. λ_t controls for seasonal fixed effect. I do not include household fixed effects, because otherwise time invariant covariates will all drop from the regressions.

The sign and significance level of β^m will provide an estimate for household level self-selection into Type m households. For example, β^1 will suggest what kind of households have only husbands participate in NREGA, and their relative household characteristics compared to other types of households. Estimates of β^m close to zero would provide that Type m households are not significantly different from other types of households, and therefore eliminating the concern on self-selection. On the other hand, if β^m is different from zero both economically and statistically, then the estimation of wage effects needs to adjust for this probability. One way of adjusting this probability is simply including the estimated probability from Equation 2.5 into the right hand side of the original model 2.3. The other way is explicitly modeling households' self-selection into NREGA, and modeling the joint decision making within the household by allowing correlated error terms[6].

The second analysis provides an alternative way of estimating labor market effects of participating NREGA. The original model in the paper has only one treatment variable, either individual own participation in NREGA or the partner's participation, in the right hand side of the equation. Such a model will yield biased estimates for the treatment effect if the spouses make a joint decision on who works in NREGA. Equation 2.9 mitigates this concern by including three treatment variables to the right hand side of the model; that is, an indicator for individual own participation in NREGA (D_{it}), an indicator for the partner's

[6]The following few lines provide an initial setup of the model.

Let D_{it}^j be an indicator for participating NREGA, where j indicates the husband or wife.

$$D_{it}^j = 1[z_{it}^j \gamma^j + \mu^j + \epsilon_{it}^j] \quad for \quad j \in \{H, W\} \tag{2.6}$$

Then individual i makes decision on labor supply in the private labor market. $i's$ employment outcome is determined by the following two latent variables:

$$V_{it}^E = x_{it}^E \beta_E + \eta_i^E + \theta_{it}^E \tag{2.7}$$

$$V_{it}^{NE} = x_{it}^{NE} \beta_{NE} + \eta_i^{NE} + \theta_{it}^{NE} \tag{2.8}$$

Individual $i's$ employment outcome is determined by $L_{it}^* = \max_l \{V_{it}^l\} forl \in \{E, NE\}$

participation (D_{it}^S) and the interaction term of these two ($D_{it}D_{it}^S$).

$$y_{it} = \beta_0 D_{it} + \delta_0 D_{it}^S + \theta_0 D_{it} D_{it}^S + \alpha_i + \lambda_t + \gamma_1 X_{it} + \gamma_2 X_{ht} + \varepsilon_{it} \tag{2.9}$$

where $D_{it} = 1$ if individual i works in NREGA and 0 otherwise; $D_{it}^S = 1$ if individual $i's$ spouse works in NREGA and 0 otherwise; y_{it} is individual $i's$ labor market outcomes including wage and workdays in the private agricultural market. X_{it} and X_{ht} are respectively individual and household demographic information. This model will give estimates for individual own response to participation in NREGA and spousal response to the partner's participation in NREGA at the same time. In particular, when restricting the sample to wives, the model will provide estimates for how wives' private labor market outcomes are affected by their own participation in NREGA (β_0) and by husbands' participation in NREGA (δ_0). Likewise, when restricting the sample to husbands, the model will provide husbands' response to their own and wives' participation in NREGA. Moreover, this model also provides the estimated effect of both spouses working in NREGA on individual private labor market outcomes.

A disadvantage of Equation 2.9 to the original model is that this model does not estimate pre-treatment trends. Therefore, an improvement to this model is, as before, using three seasons in 2006 as pre-treatment periods and estimate the following model:

$$\begin{aligned} y_{it}^p = & \left(\sum_{k=-2,-1,0} \beta_k^p D_{it}^k \right) + \left(\sum_{k=-2,-1,0} \delta_k^p D_{it}^{Sk} \right) + \left(\sum_{k=-2,-1,0} \theta_k^p D_{it}^k D_{it}^{Sk} \right) \\ & + \alpha_i^p + \lambda_t^p + \gamma_1^p X_{it}^p + \gamma_2^p X_{ht}^p + \varepsilon_{it}^p \end{aligned} \tag{2.10}$$

where $k = -2$ stands for Rabi season in 2006; $k = -1$ stands for Summer season in 2006; $k = 0$ stands for the current season. This model is essentially a combination of Equation 2.9 and Equation 2.3. $D_{it}^k = 1$ for NREGA participant

$i's$ k^{th} season from the treatment time p; $D_{it}^{Sk} = 1$ for individual $i's$ k^{th} season from the treatment time p, if $i's$ spouse participates in NREGA in season p. β_0^p gives the estimated effect of individual $i's$ participation in NREGA on own labor market outcomes; δ_0^p gives the estimated effect of the spouse's participation in NREGA on individual $i's$ labor market outcomes; θ_0^p gives the estimated response of individual $i's$ labor market outcomes to both spouses' participation in NREGA. The coefficients with subscripts -2 and -1 provide the according tests for pre-treatment trends.

Furthermore, as in the estimation equation 2.4, we can also pool the three treatment seasons in 2008 and estimate a general model. If we are not interested in heterogeneous effects by seasons, then we can estimate the following model:

$$\begin{aligned} y_{it} = & \left(\sum_{k=-2,-1,0} \beta_k D_{it}^k \right) + \left(\sum_{k=-2,-1,0} \delta_k D_{it}^{Sk} \right) + \left(\sum_{k=-2,-1,0} \theta_k D_{it}^k D_{it}^{Sk} \right) \\ & + \alpha_i + \lambda_t + \gamma_1 X_{it} + \gamma_2 X_{ht} + \varepsilon_{it} \end{aligned} \tag{2.11}$$

which will give the average effects of individual i and/or its spouse's participation in NREGA on individual $i's$ labor market outcomes.

CHAPTER 3

MINIMUM WAGE COMPETITION BETWEEN LOCAL GOVERNMENTS IN CHINA

Ravi Kanbur (Cornell University)

Yanan Li (Cornell University)

Carl Lin (Bucknell University)

3.1 Abstract

The theory of fiscal and regulatory competition between jurisdictions is more advanced than its empirical testing. This is particularly true of labor regulation in general, and minimum wage regulation in particular, and especially so for developing countries. This paper utilizes the spatial lag methodology to study city-level strategic interactions in setting and enforcing minimum wage standards during 2004-2012 in China. We manually collect a panel data set of city-level minimum wage standards from China's government websites. This analysis finds strong evidence of spatial interdependence in minimum wage standards and enforcement among main cities in China. If other cities decrease minimum wage standards by 1 RMB, the host city will decrease its standard by about 0.7-3.2 RMB. If the violation rate in other cities increases by 1 percentage point, the host city will respond by an increase of roughly 0.4-1.0 percentage points. These interactions suggest the need for policy coordination in labor regulation in China.

3.2 Introduction

Strategic interactions of fiscal policies among governments have been well discussed in both theoretical and empirical studies. Early examples of theory papers on tax competition include Kanbur and Keen (1993); Edwards and Keen (1996); Wilson and Wildasin (2004).[1] Compared to tax competition and environmental policy competition, however, jurisdictional interactions of labor standards and regulatory policies have not been studied as intensively (some examples include Duanmu, 2014; Davies and Vadlamannati, 2013).

A conventional wisdom is that there is a potential "race to the bottom" in labor standards across countries. Governments might undercut each other's labor standards to attract foreign capital (e.g. Chau and Kanbur, 2006; Davies and Vadlamannati, 2013; Olney, 2013). On the other hand, strategic interactions among jurisdictions could also lead to a "race to the top" of labor standards, for example, in the case that labor becomes a scarce resource. Regardless it is race to the bottom or to the top, the key idea is policies in one country might be influenced by those in others. While the existing literature has provided evidence on between-country interactions in labor standards, there is so far little evidence on within-country competition, especially in developing countries. Given the importance of this issue, this paper fills the gap by providing the evidence for strategic interactions on minimum wage standards in China.

This paper focuses on minimum wages as a leading example of a labor standard for the following reasons. First, minimum wage standards directly reflect

[1]Examples of other papers on tax competition, environmental regulation competition and welfare competition include: Allers and Elhorst (2005); Brueckner and Saavedra (2001); Edmark and Ågren (2008); Fredriksson and Millimet (2002); Konisky (2007); Markusen et al. (1995); Plümper et al. (2009).

the strictness of local labor market and other labor standards. For instance, an increase of minimum wages leads to a rightward shift of the whole wage distribution (Neumark et al., 2004). Second, the frequent adjustment of minimum wage standards since 2004, coupled with large spatial variations of both minimum wage standards and its enforcement, make China an ideal policy setting to identify interjurisdictional competition within a country. Before the year 2004, a minimum wage standard was close to nonexistent, with low statutory levels and weak enforcement of the laws. From the mid-2000s onwards, however, rising concerns on inequality led to considerable strengthening of regulation and enforcement. Third, minimum wages and its enforcement have a relatively decentralized decision-making system in China, which is also critical to the study of jurisdictional competition. Therefore, we use minimum wages as an example to study labor standard competition.

This paper relies on a Spatial Lag framework combined with exogenous covariates to identify the spatial interdependence in setting up and enforcing minimum wage standards in China. In the analysis of minimum wage standards, we collect a panel data set of city-level minimum wage standards from China's government websites during 2004-2012. To estimate the magnitude of spatial interdependence, we first estimate a spatial static panel data model using both Maximum Likelihood method and Instrumental Variables (IV/GMM) method, then estimate a dynamic panel data model using Arellano-Bond GMM estimator. The analysis finds strong evidence of spatial correlation in minimum wage standards. If other cities increase (or decrease) minimum wage standards by 1 RMB, the host city will increase (or decrease) its standard by about 0.7-3.2 RMB.

Then we conduct a parallel analysis of strategic interactions on minimum wage enforcement. The literature on fiscal competition has found that enforcement policies are used as instruments for fiscal competition, when competition in tax rates are banned (Cremer and Gahvari, 2000). For instance, Ronconi (2012) finds that governments react to the competitive pressures produced by FDI inflow by turning a blind eye to noncompliance of labor laws. Bhorat et al. (2012) is an example of a recent literature documenting non-compliance with minimum wage regulation in developing countries. Despite data limitations, the current paper makes the first attempt to assess competition on the enforcement of minimum wage standards. We find that if the violation rate in other cities increases by 1 percentage point, the host city will respond by an increase of roughly 0.3 percentage points.

The organization of the paper is as follows. Section 2 introduces the institutional background of minimum wage setting in China. Section 3, data. Section 4 sets the empirical model and discusses identification strategies. Section 5 presents the main results, followed by checks in Section 6. Section 7 concludes by putting our results in the context of the broader literature and looks ahead to areas for further research.

3.3 Institutional Background

3.3.1 Minimum wage setting and enforcement in China

Minimum wage regulations have been existing in China since the 1990s, but only with low level of standards and weak enforcement. Things did not change until the early 2000s, when rising income inequality became a national concern.

In 2004, the Ministry of Labor and Social Security issued a "Minimum Wage Regulations" law, stipulating that provinces should adjust minimum wage levels at least every two years to fit local living standards. Then the next decade saw frequent upward minimum wage adjustments along with improved compliance. Table 1 shows that over 60 percent of cities adjust minimum wage standards each year during 2004-2012 (except for 2009). While real minimum wage rates almost doubled during 2004-2009, the average noncompliance rate decreased from 10 to 7 percent (Table B.1 in the Appendix).

The decision process of minimum wage adjustment varies by province. In some provinces, prefecture-cities are actively involved in adjusting minimum wage standards.[2] For instance, the city Chengdu in Sichuan province and the city Shenzhen in Guangdong province could set their own minimum wage standards, according to our survey with local practitioners. In other provinces, cities and counties are sorted into several tiers based on their economic development levels (usually four). The provincial government then consults with labor unions and sets the floors of minimum wage standards for each group of cities. The wiggle room left for city governments in deciding the final minimum wage standards is articulated in two ways: which group to be grouped into; and whether (and of what magnitude) it is possible to further adjust upward the standard given the floor. For instance, the city Haimen in Jiangsu province was in tier 2 areas before 2007, but in tier 1 since 2008. Therefore, cities in such provinces have some degree of flexibility in the final decision of minimum wage standards. Overall, cities have some flexibility in setting up minimum wage standards in most provinces, which is evidenced by large within province vari-

[2]China's administrative structure includes, from high to low levels, provinces, prefecture-level cities, and counties. A prefecture-level city comprises a central urban area and several counties.

ations of minimum wage standards.

When it comes to enforcing minimum wage laws, the "Minimum Wage Regulation" law specifies that county-above governments are in charge of the enforcement in local areas. Therefore, prefecture-level cities, which stand one level above counties, will have full control over minimum wage enforcement in their cities.

3.3.2 Motivation of local leaders to compete on minimum wages

Leaders of city governments have at least the following three motivations to compete each other on minimum wage standards and enforcement. While they are all plausible and have been documented in the literature, this paper does not discriminate one source from another.

The first incentive for interjurisdictional competition on minimum wages is competing for capital, which would lead to a race to the bottom of minimum wages. If we assume capital flows to places with lower labor cost, then more stringent labor standards (e.g. minimum wages, labor rights) in local areas compared to other areas would reduce the attractiveness of local environments for firms. Therefore, local leaders might undercut each other's labor standards and employment protections to attract foreign (and domestic) capital (e.g. Chau and Kanbur, 2006; Davies and Vadlamannati, 2013; Olney, 2013).

The second incentive for inter-jurisdictional competition on minimum wages is competing for labor which, to the opposite of the first incentive, might lead to a race to the top of minimum wages. It has been argued that China has passed the Lewis turning point and the era of surplus labor is over (Zhang et al., 2011). Moreover, the last decade has seen rising labor costs across the country, because a tightening labor market in an era of high economic growth gives workers stronger wage bargaining power. Therefore, city governments might engage in the competition for labor.

Last but not least, city leaders might strategically set and enforce minimum wage standards driven by promotion incentives. Promotion of local leaders in China is decided by their upper-level governments, based on performance comparison across jurisdictions. It has been shown that such a promotion scheme leads to tournament competition among local leaders in multiple aspects, such as in investment (Yu et al., 2016) and coal mine safety (Shi and Xi, 2018). In a similar vein, minimum wage standards have been used as an important tool to curb the rising inequality; therefore, city leaders might be motivated to engage in tournament competition in minimum wage standards to appeal to the upper-level governments.

3.4 Data

We construct a panel data set from several sources. City level minimum wage standards in 2004-2012 are manually collected from local government websites (through searching "Baidu", a Chinese version of Google).[3] City characteristics and boundary shapefiles are compiled from China Data Online (see Table B.1 in

[3] In the case that there are multiple levels of minimum wage standards within one prefecture-level city, we use the highest one as that prefecture-level city's minimum wage level. In (rare) cases of two upward adjustments happening within one year (e.g. Hebei Province and Beijing city in 2007), we take the second (and higher) one as that year's minimum wage standard.

the appendix for descriptive statistics of city characteristics).

The final data set for the analysis of minimum wage standards includes 252 prefecture-level cities in 25 provincial-level administrative units as opposed to all 294 prefecture-level cities in 34 provincial-level units of China. This subsample property will not bias our estimations for the following three reasons. First, the final data set is not derived by a non-random selection process, rather it is a result of combining data sets from different sources.[4] Second, results are not driven by a particular province, as the estimates are not sensitive to dropping any individual provinces (Section 6.1 will give more details). Third, we also run two-sample t-tests for each economic variable in the omitted sample and the included sample and find that none of the mean differences is statistically significant.[5]

As discussed earlier, the frequent upward minimum wage adjustments provide a good source of identification for inter-jurisdictional dependence. Table 1 shows that about 58 percent of prefecture-level cities adjust minimum wages in each year during 2004-2012. The only exception is 2009; no provinces adjusted minimum wages due to the financial crisis. 2006, 2010 and 2011 all saw a two-digit increase of minimum wages compared to the previous years.

In the second main analysis, "race on minimum wage enforcement", we proxy the degree of enforcement by noncompliance rates of minimum wage standards in the city, which is derived from 2002-09 Urban Household Survey

[4]China Data Online originally includes 286 cities. By combining it with minimum wage data and generating a city-level data set, we lost 25 prefecture-level cities. Later, we lose 1 city when combining with the shapefile data set and drop 1 city because it is not present in all years. As a result, we get seven "island" cities that have no neighboring cities. We drop these "islands" when constructing the contiguity weighting matrix, and thereby losing 7 additional cities. Provincial-level units absent in the sample are Gansu, Qinghai, Xinjiang, Guizhou, Hainan, Guizhou, Tibet, Hongkong and Macao.

[5]The results of t-test statistics for each economic variable is available upon request.

(UHS hereafter). UHS is a continuous, large-scale social-economic survey conducted by the National Bureau of Statistics of China to study the conditions and living standards of urban households. Survey subjects include local urban households, non-local urban households who have lived in the city for at least six months, and some rural-urban migrant households. The survey covers 16 representative provinces, but our sample only keeps the cities that were surveyed throughout all years during 2004-2009. As a result, the final data set in the analysis includes only 66 prefecture-level cities distributed in nine provinces. Again, the relatively small sample size should not bias our estimates of the spillover effect of minimum wage enforcement, given the randomness of UHS sampling and sample selection process. As a matter of fact, the economic characteristics are by and large quite similar between these 66 prefecture-level cities and those dropped from the analysis.[6]

To calculate city specific noncompliance rates, we take advantage of the employment and wage information from the questionnaire.[7] Specifically, we first count the number of workers paid below local minimum wage standards, then divide this number by the employment size in the city. As Table 3.1 shows, on average, 9.2 percent of workers are underpaid during the study period. The level of minimum wage standards is on average 38 percent of the median wage in the city.

[6]The results of t-test statistics for each economic variable is available upon request.
[7]Individual monthly wage is derived by annual income/the number of months worked in the year.

Table 3.1: Trend of minimum wage standards, 2004-2012

Year	Min wage (RMB)	Growth of min wage (%)	% of jurisdictions adjusting min wage	Violation Rate	Kaitz ratio (min wage/ median wage)
2004	366	9.3%	66%	8.9%	37%
2005	395	7.9%	35%	8.5%	35%
2006	459	16.2%	88%	9.0%	37%
2007	487	6.1%	62%	10.0%	40%
2008	509	4.4%	64%	9.8%	41%
2009	512	0.6%	0%	7.6%	38%
2010	616	20.4%	99%		
2011	682	10.7%	78%		
2012	745	9.2%	68%		
Total	493	8.9%	58%	9.2%	38%

Note: Minimum wages are deflated by the provincial level CPI, using 2002 CPI as the base. City level Violation rate is calculated as the fraction of workers paid below local minimum wages.

3.5 Empirical Specification

This paper uses the spatial lag framework to estimate city-level strategic interactions in setting and enforcing minimum wage standards.[8] For each analysis, we first estimate a spatial static panel data model using both Maximum Likelihood method[9] and Instrumental Variables (IV/GMM) method, then we estimate a

[8]Other studies using the Spatial Lag framework include, for example, Brueckner and Saavedra (2001); Davies and Vadlamannati (2013); Edmark and Ågren (2008); Konisky (2007); Ollé (2003); Olney (2013); Plümper et al. (2009); Shi and Xi (2018); Yu et al. (2016).

[9]A critique by Lyytikäinen (2012) is that the cross-sectional maximum likelihood Spatial Lag estimation and Spatial IV model overestimate the degree of interdependence in tax rates, as compared to the policy change IV estimates. However, the author finds that the inclusion of municipality fixed effects in the panel data model significantly reduces the bias (in the analysis of general property tax). By this logic, as our model already takes advantage of panel data and includes city fixed effects, we expect that any potential (upward) bias would become minimum.

dynamic panel data model using Arellano-Bond GMM estimator.

3.5.1 Race on minimum wage standards

To test whether minimum wage standards in the host city depend on minimum wage standards in other cities, we estimate the following Spatial Autoregression Regression (SAR) model,

$$MW_{i,t} = \beta_0 + \rho \sum_{j \neq i} \omega_{j,i} MW_{j,t} + \beta X_{i,t-1} + \eta_i + \sigma_t + \epsilon_{i,t} \tag{3.1}$$

where $MW_{i,t}$ is the minimum wage standard in city i in year t; $\sum_{j \neq i} \omega_{j,i} MW_{j,t}$ is the Spatial Lag, the weighted average of minimum wage standards in other jurisdictional areas. $X_{i,t-1}$ is a vector of city-level economic characteristics, including GDP, per capita GDP, industry share in total GDP, labor force participation rate, the proportion of employees in primary industry, student enrollment in secondary schools, student enrollment in primary schools, the number of large-scale enterprises, the number of beds in hospitals.[10] We take a 1-year lag of the covariate vector to enhance the case for exogeneity.[11] All values such as GDP and minimum wages are deflated by the provincial level Consumer Price Index, with 2002 as the base year. η_i controls for city fixed effects; σ_t controls for year fixed effects. ρ captures the spatial dependence of minimum wages. For descriptive statistics, see Table B.1 in the Appendix.

The key independent variable, spatial lag, is constructed using two different weighting matrices $\omega_{j,i}$ in the main results, and four additional weighting

[10]Industry share in total GDP = Secondary industry GDP/ total GDP. Labor force participation = the number of employees / population. The student enrollment in secondary and primary schools and the number of hospital beds are all standardized by population.

[11]Our results remain when using 2-year or 3-year lags, but for the sake of sample size we use 1-year lag.

matrices in the robustness checks. First, we use a contiguity matrix, where a city's neighbors are defined as prefecture cities that share borders with it. The weighting matrix is normalized so that the row sum equals to unity. If the host city i has n_i neighbors, then weights are defined as,

$$\omega_{j,i} = \begin{cases} 1/n_i, \text{ if city j is one of the } n_i \text{ neighbors of city i's} \\ 0, otherwise \end{cases} \tag{W1}$$

Second, we use an inverse distance-based weighting matrix, assuming that closer cities have stronger impacts on city i than cities farther away. The neighborhood-inverse distance matrix is as follows:

$$\omega_{j,i} = 1/d_{ij}, \text{where } d_{ij} \text{ is the distance between the centroids of city i and city j} \tag{W2}$$

The main econometric challenge to identify the magnitude of the spatial interdependence is the reflection problem (Manski, 1993). Because $MW_{i,t}$ and $MW_{j\neq i,t}$ are simultaneously determined, the spatial lag term is not orthogonal to the error term. As a result, OLS does not give consistent estimates. To deal with this endogeneity, we use both Maximum Likelihood method and Instrumental Variables (IV/GMM) method to derive the estimates (Brueckner, 2003). For example, Davies and Vadlamannati (2013) use IV/GMM estimation method; Shi and Xi (2018) use MLE method. We follow the traditional spatial IV method to instrument the spatial lag with $\sum_{j\neq i}\omega_{j,i}MW_{j,t}$, the weighted average of other cities' exogeneous variables. The identification assumption is that city $j's$ exogeneous variables only affect its own minimum wage standards but do not directly impact those in city i.

In addition to estimating a static spatial lag model in Equation (3.1), we add time-lagged minimum wage standards into the right-hand side of the model, because minimum wage standards in the current year might also depend on minimum wages in the previous year. By doing so, the model becomes a dynamic panel model and fixed effect estimators are no longer consistent. To address this issue and potential endogeneity concerns, we follow Olney (2013) and estimate the dynamic model using Arellano–Bond GMM estimator (Arellano and Bond, 1991). This method takes the first difference of the model and instruments the right-hand side differenced terms with all their lagged levels.[12] We also include additional instruments, the lagged levels of weighted averages of other cities' exogenous variables, $\sum_{j \neq i} \omega_{j,i} MW_{j,t}$, which were used as instruments in estimating Equation (3.1). The estimation equation is as follows:

$$\Delta MW_{i,t} = \alpha MW_{i,t-1} + \rho \Delta \sum_{j \neq i} \omega_{j,i} MW_{j,t} + \beta \Delta X_{i,t-1} + \Delta \delta_t + \Delta \epsilon_{i,t} \tag{3.2}$$

where $\Delta MW_{i,t}$ is the change of minimum wage standards in city i from year t-1 to year t; $\sum_{j \neq i} \omega_{j,i} MW_{j,t}$ is the change of the spatially lagged minimum wage standards in city i from year $t-1$ to year t. This method identifies a causal impact of other cities' minimum wage policies on the host city's policy.

3.5.2 Race on minimum wage enforcement

To study jurisdictional interactions on enforcement, we modify the estimation equation (3.1) by changing $MW_{i,t}$ to $E_{i,t}$, where $E_{i,t}$ is the enforcement level in

[12] The estimation results are robust to using fewer lagged levels as instruments, and robust to using the Blundell and Bond (1998) SYS-GMM method instead of difference GMM. The central idea of the instruments is that, $y_{i,t-2}$ is orthogonal to the differenced error term $\Delta \epsilon_{it}$; or $E(y_{i,t-2} \Delta \epsilon_{it}) = 0$. Stata command is xtabond2 (Roodman, 2006).

city i year t. The model is as follows,

$$E_{i,t} = \beta_0 + \rho \sum_{j \neq i} \omega_{j,i} E_{j,t} + \beta X_{i,t-1} + \gamma Kaitz_{i,t} + \eta_i + \sigma_t + \epsilon_{i,t} \tag{3.3}$$

where $Kaitz_{i,t}$ indicates the ratio of minimum wage to median wage calculated for each city. This is the only additional covariate compared to Equation (3.1); we add it because cities with higher minimum wage standards tend to have lower enforcement and higher violations. Enforcement intensity could ideally be measured by the amount of resources (e.g. inspectors) that local government invests to regulate minimum wage laws, but such data is lacking. As stronger enforcement intensity is associated with lower minimum wage non-compliance (Bhorat et al., 2012), we proxy enforcement by the headcount ratio of minimum wage violations: the number of workers receiving below minimum wages divided by the total number of working population. Again, Equation (3.3) is estimated using MLE and IV method, using the same instruments as when estimating Equation (3.1).

Because the enforcement analysis uses household survey data to compute violation rates, and the household survey is conducted in a randomly selected subsample of cities in the country, we end up having many "islands" and other cities with few contiguous neighbors. Therefore, we only use the inverse distance weighting matrix to construct spatial lags. The construction of the matrix is the same as Formula (W2).

Again, we introduce a time-lagged enforcement variable into the righthand side of Equation (3.3) and estimate the dynamic panel model using Arellano-Bond method. The first differenced model is as follows,

$$\Delta E_{i,t} = \alpha E_{i,t-1} + \rho \Delta \sum_{j \neq i} \omega_{j,i} E_{j,t} + \beta \Delta X_{i,t-1} + \Delta \delta_t + \Delta \epsilon_{i,t} \tag{3.4}$$

where $\Delta E_{i,t}$ is the change of minimum wage violation rates in city i. This model identifies a causal impact of other cities' enforcement levels on the host city's enforcement.

3.6 Main Results

3.6.1 Race on minimum wage standards

Table 3.2 presents estimation results for jurisdictional interdependence on minimum wage standards using three different estimation methods and two weighting matrices. Model 1-2 both use MLE method, with model 1 using the contiguity matrix (Formula W1) and model 2 using the inverse distance matrix (Formula W2). Likewise, Model 3-4 show IV/GMM estimation results; Model 5-6 use Arellano-Bond difference GMM estimation method, using all lagged terms as instruments in addition to weighted averages of other cities' economic characteristics. Overall, these models give consistent results that other cities' minimum wage standards have a positive effect on the host city's minimum wage standards. The magnitude of spatial dependence ranges between 0.7 and 3.2. In words, if other areas increase their minimum wage by 1 RMB, the host city will increase its minimum wage by 0.7-3.2 RMB. This magnitude is comparable to that by other studies using similar methods. For instance, Davies and Vadlamannati (2013) use the IV/GMM method to estimate spatial interaction in labor standards across countries and find point estimates between 0.2 and 2. Olney (2013) finds the point estimate of intergovernmental dependence in employment protection is between 0.18 and 0.44 when using Arellano-Bond GMM method, but as large as between 1.6 and 2.9 when using the IV method.

Across the models, a few control variables have statistically significant point estimates. First, cities with higher labor force participation will have higher minimum wages. Second, higher enrollment rate of primary school students is associated with higher minimum wages, but higher enrollment rate of secondary school students is associated with lower minimum wages. Third, a higher number of large-scale enterprises is associated with higher minimum wages, which could be explained by higher demand for labor. In addition, the estimates from the dynamic estimation equation (3.2) have positive coefficients on the lagged minimum wage term, suggesting that minimum wage standards might be persistent over time.

3.6.2 Race on enforcement

Table 3.3 presents estimation results for jurisdictional interdependence on minimum wage enforcement using different estimation methods and the inverse distance weighting matrix. Enforcement is proxied by the headcount ratio of minimum wage violation. Model 1 uses MLE method; Model 2 shows IV/GMM estimation results; Model 3 uses the Arellano-Bond difference GMM estimation method.

The results consistently show evidence of a race on enforcement, with the point estimate ranging in [0.38, 0.85] across three models. In other words, if the weighted violation rate in other cities increases by 1 percentage point, the violation rate in the host city will increase by 0.38-0.85 percentage points. The insignificant coefficient on the lagged violation rate in the dynamic model (in

Table 3.2: Results of the analysis of min wage standards, 2004-12

	MLE		IV		Arellano-Bond	
	(1) Contiguous	(2) Distance	(3) Contiguous	(4) Distance	(5) Contiguous	(6) Distance
Spatial lag of minimum wages	0.72***	0.97***	0.99***	2.33***	0.98***	3.23***
	(0.02)	(0.00)	(0.11)	(0.18)	(0.04)	(0.17)
L. min wage					0.06**	0.05
					(0.03)	(0.03)
GDP (log)	5.03	-8.65	12.73*	-0.80	20.29	-33.80
	(8.45)	(10.90)	(7.31)	(8.02)	(26.01)	(26.46)
per capita GDP (log)	-8.90	-4.42	-12.99**	-9.42	-37.47**	-33.05*
	(6.36)	(8.46)	(5.55)	(6.36)	(16.55)	(19.49)
Labor participation Rate (%)	0.79**	1.16***	0.77***	1.48***	2.64**	5.22***
	(0.34)	(0.40)	(0.30)	(0.37)	(1.07)	(1.23)
Industry share (%)	-0.39	-0.61*	-0.21	-0.23	-0.09	0.51
	(0.27)	(0.35)	(0.20)	(0.22)	(0.41)	(0.53)
# of enterprises (1000)	5.55*	8.98**	3.27	8.56**	-1.84	1.61
	(2.99)	(3.85)	(3.93)	(3.45)	(2.95)	(3.54)
# of beds in hospitals	-13.49	-25.68**	-5.46	-10.15	-8.85	34.34
	(10.28)	(11.71)	(8.75)	(9.26)	(26.11)	(25.18)
stud enroll in secondary school	-1.73***	-1.73***	-1.74***	-1.67***	-0.72	-0.29
	(0.50)	(0.54)	(0.51)	(0.47)	(1.10)	(1.38)
% of secondary employment (%)	-0.33	-0.53*	-0.14	-0.20	-0.94**	-1.17*
	(0.22)	(0.28)	(0.18)	(0.21)	(0.46)	(0.60)
stud enroll in primary school	0.95	2.51**	0.02	1.48*	-0.59	-2.67
	(0.95)	(1.14)	(0.73)	(0.79)	(1.58)	(1.93)
City FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Observations	2016	2016	2016	2016	1512	1512
R2			0.96	0.94		
Kleibergen-Paap rk Wald F stat			66.21	305.17		
Hansen J statistic (p-val)			0.750	0.245	0.822	0.773
AR (2) p-value					0.998	0.210

Model 1-2 show MLE estimation results. Column 1 uses simple contiguity matrix (equation W1); column 2 uses inverse distance matrix (Equation W2). Likewise, Model 3-4 show IV/GMM estimation results; Model 5-6 use Arellano-Bond difference GMM estimation method, using all lagged terms as instruments in addition to weighted averages of other cities' economic characteristics. Results also hold using Arellano-Bond system GMM method. Control variables are taken 1-year lags. Robust standard errors in parentheses are clustered at city level. Year 2009 is dropped from all analysis, because no minimum wage adjustments occurred in any cities during the year. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

column 3) suggests that the enforcement is not persistent over time. The share of secondary industry among total GDP and the share of employment in secondary industry are both positively associated with the violation rates of the minimum wage laws. In addition, the positive estimates of Kaitz ratio (minimum wage divided by median wages) indicate that a higher minimum-to-median wage ratio results in a higher violation rate of minimum wage laws in the city.

3.7 Robustness checks

3.7.1 Robustness to sample changes

There might be potential concerns that the evidence of spatial interdependence in minimum wage standards is driven by sample selection issues. To address this concern, we conduct the following two robustness checks. In Figure 3.1 we plot the estimated spatial interdependence after drop a provincial-level unit from the whole sample. For instance, the first point shows that the magnitude of spatial interdependence is 2.35 after excluding Beijing; the second point shows that the magnitude of spatial interdependence is about 2.4 after excluding Tianjin; and so on. These estimates are derived using the IV/GMM method, while using the MLE and Arellano-Bond estimation methods give similar results. Overall, the IV/GMM point estimates are not sensitive to dropping a particular province, with the magnitudes in the range [2.1, 2.5].

In the second robustness analysis, we include only top 10 (or 20, ..., 90) percent nearest cities in the construction of inverse distance weighting matrices,

Table 3.3: Results of the analysis of minimum wage enforcement, 2004-09

	MLE	IV/GMM	Arellano-Bond GMM
	(1)	(2)	(3)
Spatial lag (Violation Rate)	0.38**	0.60*	0.85*
	(0.16)	(0.36)	(0.48)
L. Violation Rate			-0.07
			(0.11)
GDP (log)	1.78	1.99	7.60
	(1.93)	(1.67)	(14.61)
per capita GDP (log)	-0.35	-0.31	-3.93
	(1.35)	(1.45)	(6.69)
Labor participation Rate (%)	-0.12*	-0.13*	-0.03
	(0.06)	(0.07)	(0.44)
Industry share (%)	0.15***	0.14***	0.09
	(0.06)	(0.05)	(0.16)
# of enterprises (1000)	0.02	0.07	0.37
	(0.39)	(0.45)	(1.57)
# of beds in hospitals	0.29	-0.08	-3.00
	(2.92)	(2.99)	(7.95)
student enroll in secondary school	-0.32	-0.34	-0.08
	(0.30)	(0.28)	(2.29)
% of secondary employment	0.10*	0.11**	0.11
	(0.05)	(0.05)	(0.18)
student enrollment in primary school	-0.34	-0.34	0.13
	(0.24)	(0.21)	(0.82)
Kaitz ratio (%)	0.42***	0.40***	0.36***
	(0.05)	(0.05)	(0.14)
City FE	Y	Y	Y
Year FE	Y	Y	Y
Observations	396	396	264
R2		0.48	
Kleibergen-Paap rk Wald F stat		63.56	
Hansen J P-value		0.860	0.662
AR (2) p-value			0.360

Model 1 shows MLE estimation results. Model 2 shows IV/GMM estimation results; Model 3 uses Arellano-Bond difference GMM estimation method. All models use inverse distance weighting matrices to construct the spatial lag term. Control variables are taken 1-year lags. Robust standard errors in parentheses are clustered at city level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

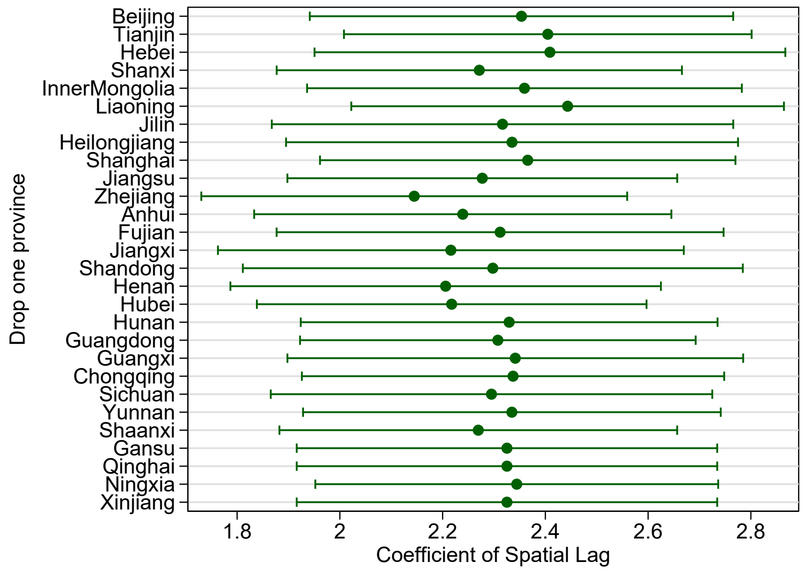


Figure 3.1: Robustness of minimum wage competition to dropping a province

Each point estimate is the coefficient of the spatial lag of minimum wages, using IV/GMM estimation method and inverse distance weighting matrix, after dropping the labelled province.

then re-estimate the model using the IV/GMM method and plot the coefficients of the spatial lag term in Figure 3.2. This subsample analysis addresses the concern that the host city might put zero weight on some faraway cities, and hence using all cities to construct the weighting matrix might be too general. To implement this idea, we first rank the distance of all other cities to the host city. Assume that the host city only references minimum wages standards in cities that are within, for example, the top 10 percent in distance, or 25 (=10% * 252) closest cities. Then we re-estimate the model using the IV/GMM method. The estimated spatial interdependence is about 1.1, as plotted in Figure 3.2. Overall, as more cities are included in the weighting matrix, the spatial interdependence gets stronger.

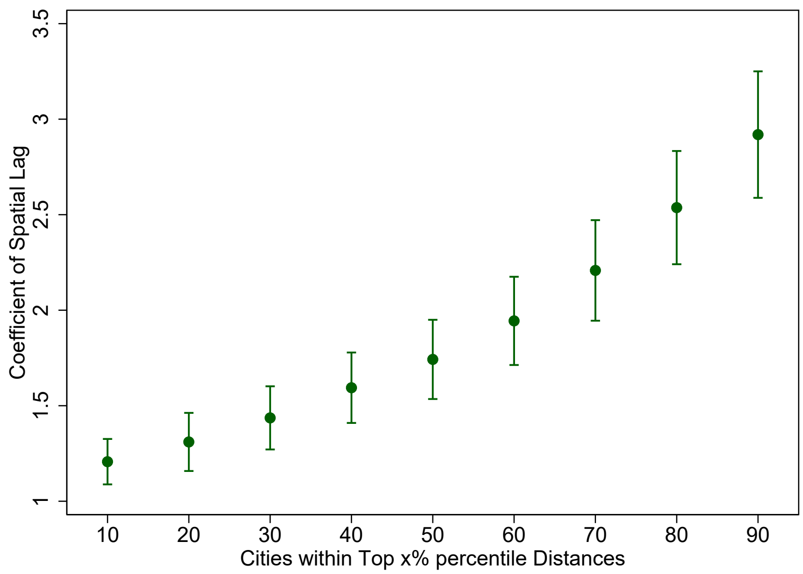


Figure 3.2: Robustness of minimum wage competition to the exclusion of distant cities

Note: Each point estimate is the coefficient of the spatial lag of minimum wages, using IV/GMM estimation method and inverse distance weighting matrix, including the top x percent nearest cities.

Likewise, there might be potential concerns that the evidence of spatial interdependence in minimum wage enforcement is also driven by sample selection. We conduct similar subsample analysis and plot the figures parallel to Figure 3.1 and 3.2. Figure 3.3 suggests that the evidence of spatial interdependence in minimum wage enforcement is robust to dropping any provinces, with the estimated magnitude ranging between 0.2 and 1.1. Figure 3.4 suggests that the evidence is also robust to the exclusion of faraway cities from the distance weighting matrix.

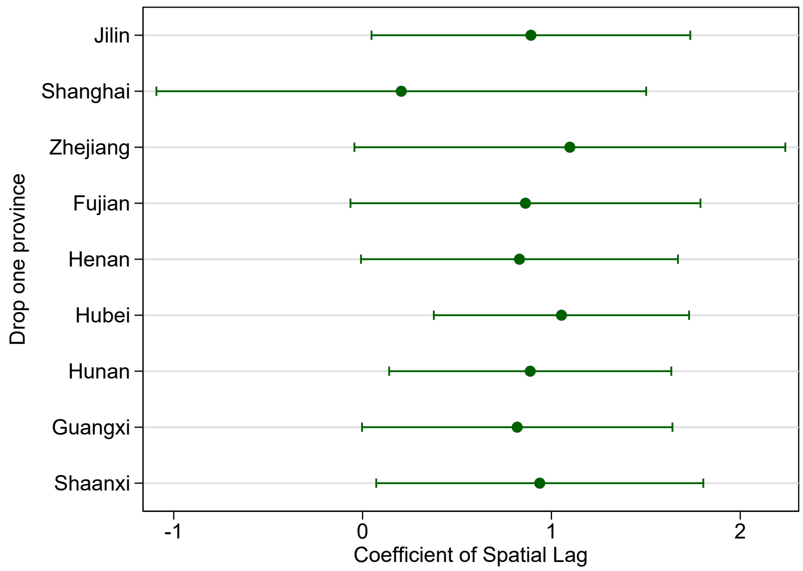


Figure 3.3: Robustness of minimum wage enforcement competition to dropping a province

Note: Each point estimate is the coefficient of the spatial lag of minimum wage violation rates, using IV/GMM estimation method and inverse distance weighting matrix, after dropping the labelled province.

3.7.2 Robustness of results to alternative weighting matrices

In the main results, we use distance-based weighting matrices to derive the estimates. However, economic distance might also be important. A city might be more likely to reference minimum wage standards in cities that have similar economic development levels rather than cities do not. Using economic characteristics as weights is common in the literature (e.g. Davies and Vadlamannati, 2013; Olney, 2013). Therefore, we construct four economic characteristics-based weighting matrices.

Economic characteristics are all taken from the year 2003, one year prior

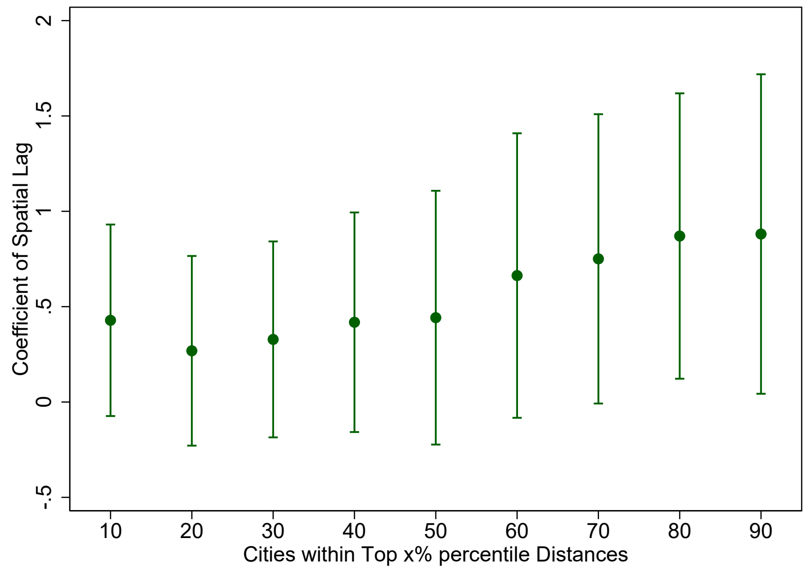


Figure 3.4: Robustness of minimum wage enforcement competition to the exclusion of distant cities

Note: Each point estimate is the coefficient of the spatial lag of minimum wage violation rates, using IV/GMM estimation method and inverse distance weighting matrix, including the top x% nearest cities.

to the first year of our analysis, so that the weighting matrix is arguably exogenous. In the case of GDP weighting matrix, the weight is given by $\omega_{j,i} = \frac{1}{|lnGDP_{j,t=2003} - lnGDP_{i,t=2003}|}$, then standardized to row sum equal to 1. Likewise, we replace log GDP by three other economic indicators, labor force participation rate, the number of enterprises and per capita GDP, to construct three other weighting matrices. Lastly, we dot multiply the inverse distance weighting matrix and the per capita GDP weighting matrix to construct a new matrix, which will then incorporate both geographical and economic distance. Table 3.4 presents estimation results using these weighting matrices and different estimation methods. Again, the results are robust to the choice of weighting matrices. All models give consistent results of spatial dependence in minimum wage standards. The magnitudes of estimation are also comparable to those in the main results.

Similarly, we construct economic characteristics-based weighting matrices and re-estimate the race on the enforcement models. Results in Table 3.5 show that MLE and Arellano-Bond estimations yield significant results, while IV/GMM estimation does not.

3.8 Conclusion

The theory of fiscal and regulatory competition between jurisdictions is more advanced than its empirical testing. This is particularly true of labor regulation in general, and minimum wage regulation in particular, and especially so for developing countries. Olney (2013) finds evidence of a race to the bottom in employment protection among OECD countries, with reaction coefficient of 1.0-2.8. Davies and Vadlamannati (2013) find labor rights in one country are positively correlated with those elsewhere, i.e. a cut in labor rights in other countries reduces labor rights in the host country, with the reaction coefficient about 0.55-0.88. They also argue that international competition lies more in enforcement than in labor laws.

This paper focuses on within-country competition on labor standards, and takes up the case of China, which introduced a vigorous minimum wage regime from the mid-2000s onwards. The analysis utilizes the spatial lag framework and three estimation methods (including maximum likelihood estimation, IV/GMM method and Arellano-Bond GMM method) to study city-level strategic interactions in setting and enforcing minimum wage standards during

Table 3.4: Robustness of minimum wage competition to alternative weighting matrices

	2003 GDP (log) (1)	Labor force part. rate (2)	# of enterprise (3)	Distance & (log) per cap GDP (4)
Panel A: MLE Results				
Spatial lag	0.16***	0.18***	0.38***	0.92***
	(0.04)	(0.05)	(0.08)	(0.02)
Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	2016	2016	2016	2016
Panel B: IV/GMM Results				
Spatial lag	1.42***	0.73***	1.12***	1.07***
	(0.30)	(0.27)	(0.28)	(0.07)
Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	2016	2016	2016	2016
R2	0.88	0.90	0.90	0.96
Kleibergen-Paap rk Wald F statistics	52.84	49.98	225.00	179.60
Hansen J statistic	0.52	0.40	0.49	0.23
Panel C: Arellano-Bond GMM Results				
Spatial lag	0.63***	0.32**	0.98***	1.19***
	(0.15)	(0.15)	(0.19)	(0.04)
L. min wage	0.10**	0.14***	0.08*	0.09***
	(0.04)	(0.05)	(0.04)	(0.03)
Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1512	1512	1512	1512
Hansen J P-value	0.737	0.747	0.736	0.781
AR (2) p-value	0.395	0.250	0.414	0.371

Each column uses a different weighting matrix to construct the spatial lag. Column (1) uses 2003 GDP (log); column (2) labor force participation rate; column (3) the number of enterprises; column (4) combines inverse distance and per capita GDP. Panel A show MLE estimation results for each model; Panel B IV/GMM estimation results; Panel C uses Arellano-Bond difference GMM estimation methods, using all lagged terms as instruments in addition to weighted averages of other cities' economic characteristics. Control variables are taken 1-year lags. Robust standard errors in parentheses are clustered at city level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

Table 3.5: Robustness of enforcement competition to alternative weighting matrices

	2003 GDP (log) (1)	Labor force part. rate (2)	# of enterprise (3)	Distance & (log) per cap GDP (4)
Panel A: MLE Results				
Spatial lag (violation rate)	0.06	0.20***	0.21*	0.13*
	(0.09)	(0.06)	(0.11)	(0.07)
Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	396	396	396	396
Panel B: IV/GMM Results				
Spatial lag (violation rate)	0.18	0.15	0.09	0.26
	(0.61)	(0.40)	(0.48)	(0.54)
Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	396	396	396	396
R2	0.46	0.48	0.47	0.47
Kleibergen-Paap rk Wald F statistics	8.33	9.18	9.64	5.34
Hansen J statistic	0.59	0.69	0.73	0.10
Panel C: Arellano-Bond GMM Results				
Spatial lag (violation rate)	0.68*	0.72**	0.70**	0.26
	(0.35)	(0.32)	(0.33)	(0.36)
L. min wage (violation rate)	-0.05	-0.16	-0.07	-0.06
	(0.12)	(0.13)	(0.08)	(0.12)
Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	264	264	264	264
Hansen J P-value	0.701	0.744	0.535	0.887
AR (2) p-value	0.419	0.461	0.138	0.114

Each column uses a different weighting matrix to construct the spatial lag. Column (1) uses 2003 GDP (log); column (2) labor force participation rate; column (3) the number of enterprises; column (4) combines inverse distance and per capita GDP. Panel A show MLE estimation results for each model; Panel B IV/GMM estimation results; Panel C uses Arellano-Bond difference GMM estimation methods, using lagged terms as instruments in addition to weighted averages of other cities' economic characteristics. Control variables are taken 1-year lags. Robust standard errors in parentheses are clustered at city level. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

2004-2012 in China. We manually collect a panel data set of city-level minimum wage standards from China's government websites. The analysis finds strong evidence of spatial correlation in minimum wage standards and enforcement among main cities in China. If other cities decrease minimum wage standards by 1 RMB, the host city will decrease its standard by about 0.7-3.2 RMB. If the violation rate in other cities increases by 1 percentage point, the host city will respond by an increase of roughly 0.4-1.0 percentage points.

The Chinese government has expanded minimum wage intervention greatly, in response to concerns about rising inequality. Our results show that there is significant interjurisdictional competition on the level of the minimum wage and in enforcement among local governments. Such competition could be wasteful, and lead to a race to the bottom, undermining the government's objectives. The interactions identified in this paper thus suggest the need for policy coordination on labor regulation in China.

Our analysis has broader significance given the resurgence of interest in minimum wages in developing countries as an instrument for addressing rising inequality. Thus Bhorat et al. (2017) provide a review of minimum wages in Africa. They find that "most countries in Sub-Saharan Africa (SSA) have adopted minimum wage regulation" and that "SSA as whole reflects a bias towards a more aggressive minimum wage policy compared to the rest of the world." In South Africa, for example, the current government has proposed a national minimum wage to replace the collection of sector and region specific minimum wages. The question of whether to allow local setting of minimum wages to take account of local conditions is an area of open debate. In Asia, the decentralization reforms in Indonesia were accompanied by a decree allowing local governments to set minimum wages. As countries like Myanmar start a new era of labor regulation, the questions of minimum wages and local flexibility in implementation are at the forefront. In Russia, minimum wage setting was decentralized in 2007. Around the world, therefore, interjurisdictional competition in minimum wages is a live issue. Our analysis provides an initial framework in which competing perspectives on these debates can be assessed quantitatively.

Our evidence on jurisdictional interdependence in minimum wage setting within a country also raises a set of interesting further research questions. What we have shown is that local government react to each other in setting minimum wages, and in enforcement of minimum wages. A natural interpretation of that is a possible "race to the bottom", as jurisdictions lower labor standards to attract investment. But could there also be a "race to the top" in other dimensions? Rather than lower labor standards, a local government could improve infrastructure, or improve the quality of local governance, to make investment more attractive. This could set in motion a chain of reactions through which other localities respond by improving their infrastructure and business environment so that there is an upward virtuous cycle of overall improvement in labor productivity rather than a downward vicious spiral of lowering labor costs through lowering labor standards. This raises the empirical question—do we see such a virtuous race to the top in practice? And the policy question—what can the government do to trigger the virtuous spiral?

APPENDIX A
CHAPTER 2 OF APPENDIX

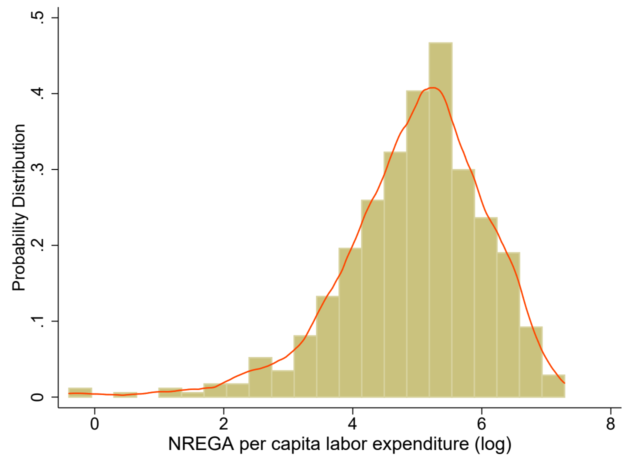


Figure A.1: Distribution of district-wise NREGA labor expenditure during 2006-2010

Note: Per capita labor expenditure is calculated as total labor expenditure of NREGA projects (Rupees) divided by rural population in the district. NREGA work data comes from MGN-REGA Public data portal. Only districts in the final regression sample are included.

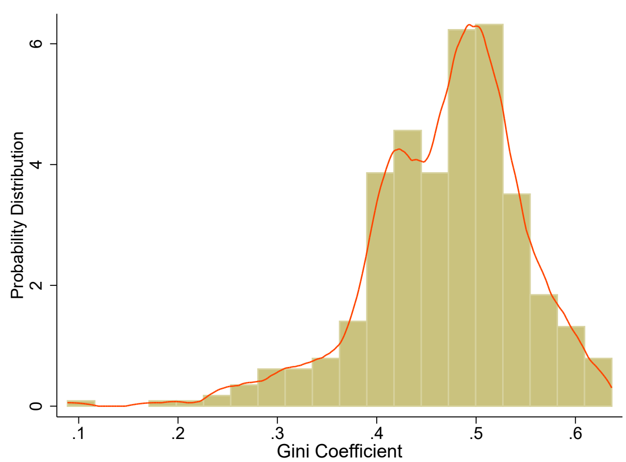


Figure A.2: Probability Distribution of Gini Coefficient, 2005

Source: The author calculated Gini coefficient of landownersihp based on district-wise land distribution data from 2005 India Agricultural census.

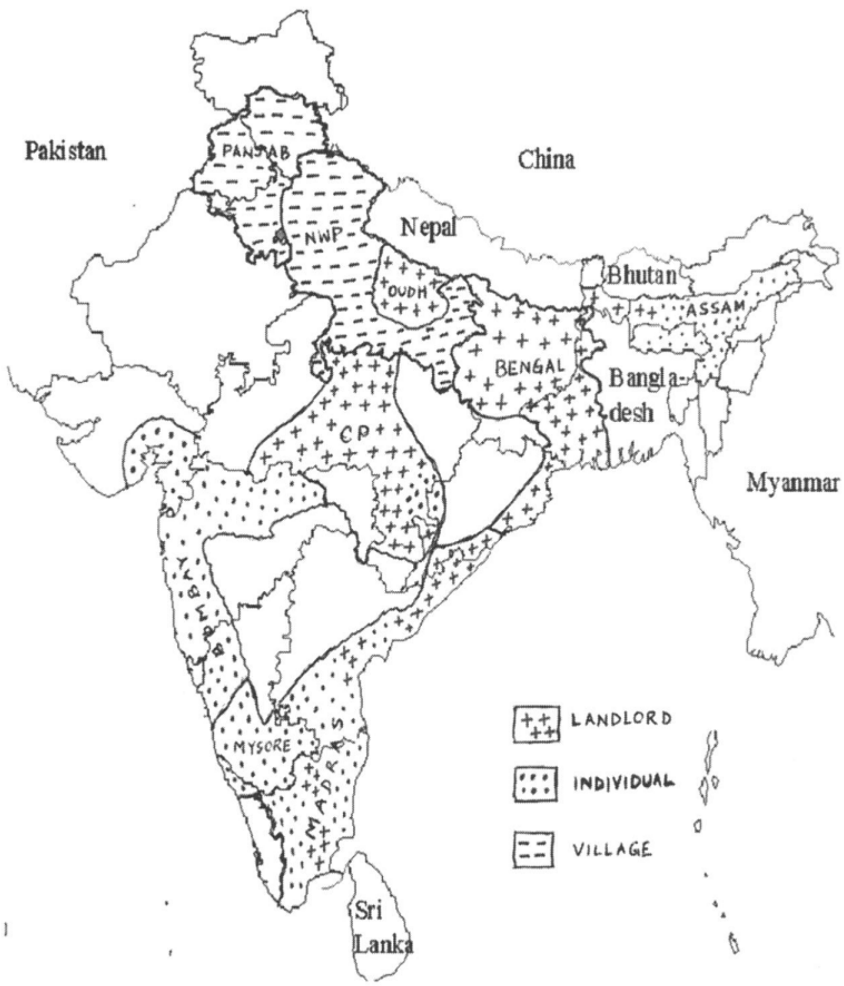


Figure A.3: Map of India

Source: Banerjee & Iyer (2005).

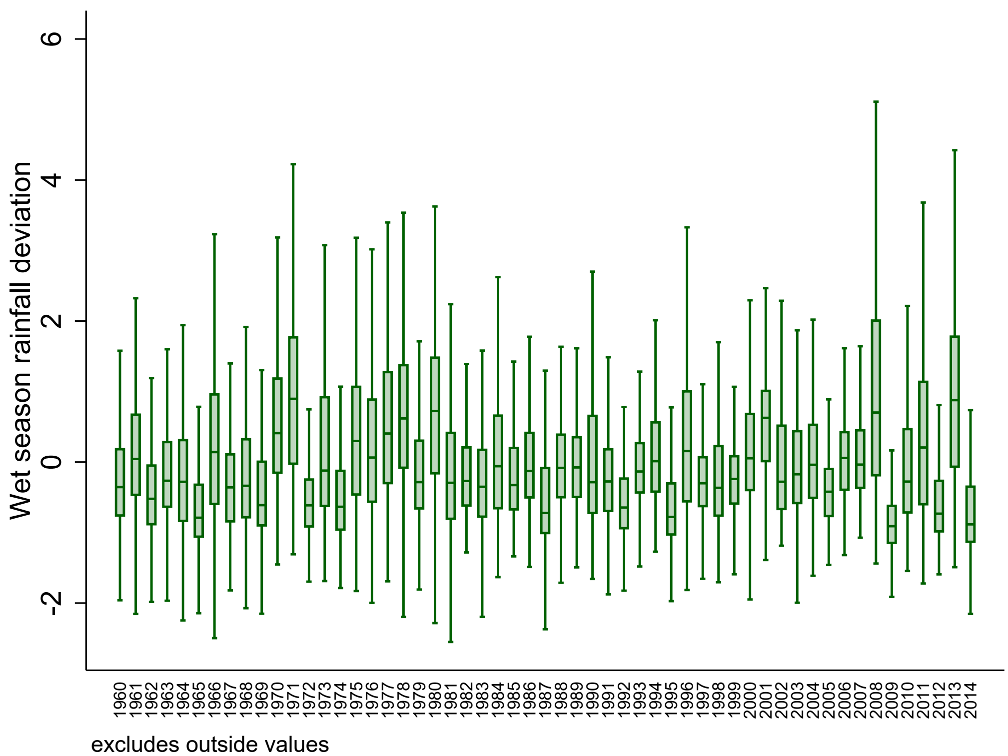


Figure A.4: Rainfall deviation during wet season, 1973-2002

Source: Original data is downloaded from India Water Portal. District-year wise rainfall deviation is the difference between rainfall during the main growing season (wet season = June-November) and the long-term average in the district, normalized by the standard deviation.

Table A.1: Comparison of Landlord and Non-Landlord district characteristics

	Mean Difference	Standard Error
Gini coef.	-0.039***	(0.010)
Rural area (Sq. km)	624.982	(665.989)
wet season rainfall (100 mm)	0.041	(0.046)
% of land covered in fine soil	-7.299***	(2.340)
% of land covered in medium soil	12.757**	(5.138)
% of Rural pop	7.522***	(1.853)
Literacy rate	-5.585***	(1.900)
% of SC population	1.506	(1.536)
% of ST population	-0.821	(1.638)
Work-population ratio	0.507	(0.852)
% of Main workers	0.676	(0.756)
% of Marginal workers	0.048	(0.845)
% of non Workers	-0.507	(0.852)
% of Agricultural labourers	-0.984	(1.660)
% of Cultivators	8.279***	(1.984)
% of Household industry workers	-1.116	(0.787)
% of Other industries	-6.303***	(1.805)
% villages with Safe Drinking water	-0.009	(0.462)
% villages with Electricity (Power Supply)	-6.726*	(3.503)
% villages with Paved approach road	-5.404**	(2.628)
% villages with Primary school	-0.960	(2.748)
% villages with Medical facility	-4.433	(3.386)
% villages with Post and telephone facility	-2.977	(3.297)
Ag Wages (Rs/day 1996)	-0.131***	(0.033)
Rs. per Hectare. 1990-1993	-0.562	(0.420)
Composite Backwardness Index	-0.100**	(0.049)
Observations	457	

This table shows the difference of geographic and economic characteristics between landlord and non-landlord districts. Each row estimate is derived by regressing the row variable on gini coefficient and state dummies. The sample is restricted at the IV sample in Table 1.7.

Table A.2: Comparison of Land distribution in 2005 and 2010

Variables	2005 (mean)	2010 (mean)	Mean Diff
top10	0.359	0.357	-0.002
top20	0.528	0.525	-0.003
top30	0.648	0.645	-0.003
top40	0.739	0.735	-0.004
bot40	0.136	0.140	0.004
bot30	0.092	0.095	0.003
bot20	0.056	0.059	0.002
bot10	0.026	0.028	0.001
mid40_80	0.380	0.380	0
mid50_80	0.336	0.336	0
Gini	0.466	0.460	-0.006
Number of districts	528	561	

Notes: Top10 means the share of land by top 10% land holdings. mid40-80 denotes the share of land by middle 40-80 percent of land holdings.

Table A.3: Dep var: Share of households provided with NREGA jobs

	(1) OLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS
Gini coef.	-0.23***	-0.76**	-0.71***	-0.70***	-0.77***
	(0.07)	(0.30)	(0.20)	(0.23)	(0.24)
log rural area(Sq. km)	0.24		0.38*	0.60***	0.42*
	(0.17)		(0.22)	(0.23)	(0.23)
log Rural population	-0.39**		-0.55***	-0.35*	-0.60***
	(0.15)		(0.20)	(0.20)	(0.21)
Literacy rate	-0.92		-0.88	0.70	-0.84
	(0.63)		(0.80)	(0.89)	(0.84)
Wet season rainfall deviation	-0.07*		-0.11***	-0.12***	-0.11***
	(0.04)		(0.04)	(0.04)	(0.04)
% of land covered in fine soil	-1.67***		-3.28***	-3.27***	-3.64***
	(0.54)		(0.80)	(0.92)	(0.94)
% of land covered in medium soil	-0.67*		-1.60***	-1.51**	-1.85***
	(0.39)		(0.53)	(0.62)	(0.59)
% of Agricultural labourers	1.91***		2.44***	3.57***	2.72***
	(0.56)		(0.73)	(0.99)	(0.81)
% of Main workers	3.90**		5.40***	8.03***	5.30***
	(1.59)		(1.93)	(2.21)	(2.00)
% of Marginal workers	-1.79		-1.37	3.45	-1.20
	(1.50)		(1.80)	(2.14)	(1.89)
% of SCST population	1.05**		0.56	1.91***	0.76
	(0.47)		(0.61)	(0.71)	(0.62)
% villages with Safe Drinking water	-1.02		-2.07	-1.63	-1.88
	(3.18)		(3.29)	(3.93)	(3.44)
% villages with Electricity (Power Supply)	1.06**		1.03*	0.36	0.92
	(0.42)		(0.57)	(0.65)	(0.59)
% villages with Paved approach road	0.37		1.29**	0.54	1.36**
	(0.44)		(0.61)	(0.64)	(0.64)
% villages with Primary school	1.31***		1.60***	2.17***	1.58***
	(0.46)		(0.49)	(0.53)	(0.50)
% villages with Medical facility	0.45		0.55*	0.60	0.53
	(0.31)		(0.33)	(0.44)	(0.34)
% villages with Post and telephone facility	-1.93***		-1.96***	-2.33***	-1.94***
	(0.38)		(0.40)	(0.49)	(0.41)
Phase 2 indicator	-0.77***		-0.79***		-0.80***
	(0.10)		(0.13)		(0.13)
Phase 3 indicator	-0.99***		-0.97***		-1.03***
	(0.11)		(0.13)		(0.14)
Composite Backwardness Index	-0.00			0.13	0.37*
	(0.11)			(0.23)	(0.19)
Observations	570	570	570	570	570
First-stage F statistics		15.08	17.77	14.83	14.73

Notes: Column 1 shows OLS results; Column 2-5 present IV estimates with different sets of covariates added. Dependent variable is the share of households provided with NREGA jobs in each year during 2006-2010. District-wise land Gini coefficient is constructed using 2005 Indian Agricultural census. They are standardized by respective standard deviation of the sample. Instrumental variable is a binary indicator that equals 1 if the district in question was a landlord district (i.e. landlords were responsible for collecting land revenue collection) in British Raj. All models include year and state dummy. * $p < 0.10$, ** $p<0.05$, *** $p<0.01$.

APPENDIX B

CHAPTER 3 OF APPENDIX

Table B.1: Descriptive Statistics, 2004-2012

	Obs	Mean	Std. Dev	Min	Max
GDP (log)	2,268	5.7	1.2	2.5	9.9
per capita GDP (log)	2,268	10.3	0.7	8.0	12.6
Labor participation Rate (%)	2,268	17.6	10.4	1.8	97.4
Industry share (%)	2,268	51.5	12.0	8.0	89.0
# of enterprises (1000)	2,268	0.6	1.4	0.0	18.5
# of beds in hospitals (per 100 people)	2,268	0.5	0.2	0.0	2.5
stud enroll in secondary sch (per 100 people)	2,268	6.6	2.6	0.9	71.6
% of secondary employment	2,268	47.6	14.0	3.8	81.9
stud enroll in primary sch (per 100 people)	2,268	7.8	3.0	1.1	32.5
Minimum wages (deflated by 2002 CPI)	2,268	530.3	155.9	212.0	1164.2
Kaitz ratio (%)	396	38.0	9.1	18.6	72.0

BIBLIOGRAPHY

Agriculture Census Report (2005). All india report on agriculture census 2005-06. Technical report, Ministry of Agriculture, Department of Agriculture & Cooperation.

Alesina, A. and Rodrik, D. (1994). Distributive politics and economic growth. *The Quarterly Journal of Economics*, 109(2):465–490.

Allers, M. A. and Elhorst, J. P. (2005). Tax Mimicking and Yardstick Competition Among Local Governments in the Netherlands. *International Tax and Public Finance*, 12(4):493–513.

Anderson, S., Francois, P., and Kotwal, A. (2015). Clientelism in indian villages. *The American Economic Review*, 105(6):1780–1816.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297.

Azam, M. (2011). The impact of indian job guarantee scheme on labor market outcomes: Evidence from a natural experiment. *Available at SSRN 1941959*.

Ballot, G., Fakhfakh, F., and Taymaz, E. (2006). Who benefits from training and r&d, the firm or the workers? *British journal of industrial relations*, 44(3):473–495.

Banerjee, A. and Iyer, L. (2005). History, institutions, and economic performance: the legacy of colonial land tenure systems in india. *The American Economic Review*, 95(4):1190–1213.

Bardhan, P. and Mookherjee, D. (2005). Decentralizing antipoverty program delivery in developing countries. *Journal of Public Economics*, 89(4):675–704.

Basu, A. K., Chau, N. H., and Kanbur, R. (2009). A theory of employment guarantees: Contestability, credibility and distributional concerns. *Journal of Public economics*, 93(3):482–497.

Basu, P. and Sen, K. (2015). Welfare implications of india's employment guarantee programme with a wage payment delay.

Baten, J. and Juif, D. (2014). A story of large landowners and math skills: Inequality and human capital formation in long-run development, 1820–2000. *Journal of Comparative Economics*, 42(2):375–401.

Beg, S. (2016). Traditional elites: Agricultural productivity and the persistence of political power. *Working paper*.

Benabou, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review*, pages 96–129.

Berg, E., Bhattacharyya, S., Rajasekhar, D., Manjula, R., et al. (2014). Can public employment schemes increase equilibrium wages? evidence from a natural experiment in india. Technical report, Department of Economics, University of Bristol, UK.

Besley, T. and Burgess, R. (2000). Land reform, poverty reduction, and growth: Evidence from india. *The Quarterly Journal of Economics*, 115(2):389–430.

Besley, T., Leight, J., Pande, R., and Rao, V. (2016). Long-run impacts of land regulation: Evidence from tenancy reform in india. *Journal of Development Economics*, 118:72–87.

Bhorat, H., Kanbur, R., and Mayet, N. (2012). Estimating the Causal Effect of Enforcement on Minimum Wage Compliance: The Case of S outh A frica. *Review of Development Economics*, 16(4):608–623.

Bhorat, H., Kanbur, R., and Stanwix, B. (2017). Minimum Wages in Sub-Saharan Africa: A Primer. *The World Bank Research Observer*, 32(1):21–74.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

Bose, N. and Das, S. (2015). Political reservation for women and delivery of public works program.

Boustan, L., Ferreira, F., Winkler, H., and Zolt, E. M. (2013). The effect of rising income inequality on taxation and public expenditures: Evidence from us municipalities and school districts, 1970–2000. *Review of Economics and Statistics*, 95(4):1291–1302.

Brueckner, J. K. (2003). Strategic Interaction Among Governments: An Overview of Empirical Studies. *International Regional Science Review*, 26(2):175–188.

Brueckner, J. K. and Saavedra, L. A. (2001). Do Local Governments Engage in Strategic Property-Tax Competition? *National Tax Journal*, 54(2):203–230.

Card, D., Devicienti, F., and Maida, A. (2014). Rent-sharing, holdup, and wages: Evidence from matched panel data. *The Review of Economic Studies*, 81(1):84–111.

Census (2011). Administrative atlas of india - census of india 2011. Technical report, http://censusindia.gov.in/2011census/maps/maps2011.html.

Chau, N. and Kanbur, R. (2006). The Race to the Bottom, from the Bottom. *Economica*, 73(290):193–228.

Cinnirella, F. and Hornung, E. (2016). Landownership concentration and the expansion of education. *Journal of Development Economics*, 121:135–152.

Commission, P. et al. (2003). Identification of districts for wage and self employment programmes. *Report of the Task Force*.

Conley, T. G., Hansen, C. B., and Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272.

Cremer, H. and Gahvari, F. (2000). Tax evasion, fiscal competition and economic integration. *European Economic Review*, 44(9):1633–1657.

Dasgupta, A., Gawande, K., and Kapur, D. (2017). (when) do antipoverty programs reduce violence? india's rural employment guarantee and maoist conflict. *International Organization*, pages 1–28.

Davies, R. B. and Vadlamannati, K. C. (2013). A race to the bottom in labor standards? An empirical investigation. *Journal of Development Economics*, 103:1–14.

Del Ninno, C., Subbarao, K., Milazzo, A., et al. (2009). *How to make public works work: A review of the experiences*. World Bank, Human Development Network.

Devereux, S. and Solomon, C. (2006). *Employment creation programmes: The international experience*. ILO.

Ding, W., Lehrer, S. F., Rosenquist, J. N., and Audrain-McGovern, J. (2009). The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics*, 28(3):578–597.

Duanmu, J.-L. (2014). A race to lower standards? Labor standards and location choice of outward FDI from the BRIC countries. *International Business Review*, 23(3):620–634.

Dunning, T. and Nilekani, J. (2013). Ethnic quotas and political mobilization: caste, parties, and distribution in indian village councils. *American Political Science Review*, 107(01):35–56.

Dutta, P., Murgai, R., Ravallion, M., and Van de Walle, D. (2014). Does india's employment guarantee scheme guarantee employment?

Dutta, P., Murgai, R., Ravallion, M., and Van de Walle, D. P. (2012). Does india's employment guarantee scheme guarantee employment? *World Bank Policy Research Working Paper*, (6003).

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics*, 84(2):755–776.

Edmark, K. and Ågren, H. (2008). Identifying strategic interactions in Swedish local income tax policies. *Journal of Urban Economics*, 63(3):849–857.

Edwards, J. and Keen, M. (1996). Tax competition and Leviathan. *European Economic Review*, 40(1):113–134.

Fishe, R. P. (1982). Unemployment insurance and the reservation wage of the unemployed. *The Review of Economics and Statistics*, pages 12–17.

Fredriksson, P. G. and Millimet, D. L. (2002). Strategic Interaction and the Determination of Environmental Policy across U.S. States. *Journal of Urban Economics*, 51(1):101–122.

Gaiha, R. (1996). Wages, participation and targeting—the case of the employment guarantee scheme in india. *Journal of International Development*, 8(6):785–803.

Galor, O., Moav, O., and Vollrath, D. (2009). Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. *The Review of Economic Studies*, 76(1):143–179.

Gulzar, S. and Pasquale, B. (2016). Politicians, bureaucrats, and development: Evidence from india. *American Political Science Review*, pages 1–22.

Gupta, B. and Mukhopadhyay, A. (2016). Local funds and political competition: evidence from the national rural employment guarantee scheme in india. *European Journal of Political Economy*, 41:14–30.

Imbert, C. and Papp, J. (2015). Labor market effects of social programs: Evidence from india's employment guarantee. *American Economic Journal: Applied Economics*, 7(2):233–263.

Jacobson, L. S., LaLonde, R. J., and Sullivan, D. G. (1993). Earnings losses of displaced workers. *The American economic review*, pages 685–709.

Jha, R., Gaiha, R., and Shankar, S. (2008). Reviewing the national rural employment guarantee programme. *Economic and Political Weekly*, pages 44–48.

Kanbur, R. and Keen, M. (1993). Jeux Sans Frontieres: Tax Competition and Tax Coordination When Countries Differ in Size. *American Economic Review*, 83(4):877–92.

Kiefer, N. M. and Neumann, G. R. (1979). An empirical job-search model, with a test of the constant reservation-wage hypothesis. *The Journal of Political Economy*, pages 89–107.

Konisky, D. M. (2007). Regulatory Competition and Environmental Enforcement: Is There a Race to the Bottom? *American Journal of Political Science*, 51(4):853–872.

López-Calva, L. F. et al. (2017). World development report 2017: Governance and the law.

Lyytikäinen, T. (2012). Tax competition among local governments: Evidence from a property tax reform in Finland. *Journal of Public Economics*, 96(7-8):584–595.

Maiorano, D. (2014). The politics of the mahatma gandhi national rural employment guarantee act in andhra pradesh. *World Development*, 58:95–105.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542.

Markusen, J., Morey, E., and Olewiler, N. (1995). Competition in regional environmental policies when plant locations are endogenous. *Journal of Public Economics*, 56(1):55–77.

McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, pages 113–126.

Muralidharan, K., Niehaus, P., and Sukhtankar, S. (2016). General equilibrium effects of (improving) public employment programs: Experimental evidence from india. *Department of Economics, University of California, San Diego, processed*.

Nath, A. (2015). Bureaucrats and politicians: How does electoral competition affect bureaucratic performance? *Institute for Economic Development (IED) Working Paper*, 269.

Neumark, D., Schweitzer, M., and Wascher, W. (2004). Minimum Wage Effects throughout the Wage Distribution. *The Journal of Human Resources*, 39(2):425–450.

Niehaus, P. and Sukhtankar, S. (2013). The marginal rate of corruption in public programs: Evidence from india. *Journal of Public Economics*, 104:52–64.

NREGA Operational Guidelines (2013). Mahatma gandhi national rural employment gurantee act 2005 – operational guidelines, 4th version. Technical report, Ministry of Rural Development, Department of Rural Development.

NREGA Report (2007). Mahatma gandhi national rural employment gurantee act 2005 – report of the second year april 2006 – march 2007. Technical report, Ministry of Rural Development, Department of Rural Development.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in africa. *The American Economic Review*, 101(7):3221–3252.

Ollé, A. S. (2003). Electoral accountability and tax mimicking: the effects of electoral margins, coalition government, and ideology. *European Journal of Political Economy*, 19(4):685–713.

Olney, W. (2013). A race to the bottom? Employment protection and foreign direct investment. *Journal of International Economics*, 91(2):191–203.

Persson, T. and Tabellini, G. (1994). Is inequality harmful for growth? *The American Economic Review*, pages 600–621.

Plümper, T., Troeger, V. E., and Winner, H. (2009). Why is There No Race to the Bottom in Capital Taxation? *International Studies Quarterly*, 53(3):761–786.

Ramcharan, R. (2010). Inequality and redistribution: Evidence from us counties and states, 1890–1930. *The Review of Economics and Statistics*, 92(4):729–744.

Ravi, S. and Engler, M. (2015). Workfare as an effective way to fight poverty: the case of India's NREGS. *World Development*, 67:57–71.

Ronconi, L. (2012). Globalization, Domestic Institutions, and Enforcement of Labor Law: Evidence from Latin America. *Industrial Relations: A Journal of Economy and Society*, 51(1):89–105.

Roodman, D. (2006). How to do xtabond2: An introduction to difference and system gmm in stata.

Santangelo, G. (2016). Firms and farms: The impact of agricultural productivity on the local indian economy. *Working Paper*.

Shah, M. and Steinberg, B. M. (2015). Workfare and human capital investment: Evidence from india. Technical report, National Bureau of Economic Research.

Sheahan, M., Liu, Y., Barrett, C. B., and Narayanan, S. (2016). Preferential resource spending under an employment guarantee: the political economy of mgnregs in andhra pradesh. *The World Bank Economic Review*, page 044.

Shi, X. and Xi, T. (2018). Race to safety: Political competition, neighborhood effects, and coal mine deaths in China. *Journal of Development Economics*, 131(C):79–95.

Sokoloff, K. L. and Engerman, S. L. (2000). History lessons: Institutions, factors endowments, and paths of development in the new world. *The Journal of Economic Perspectives*, 14(3):217–232.

Subbarao, K. (1997). Public works as an anti-poverty program: An overview of cross-country experience. *American Journal of Agricultural Economics*, 79(2):678–683.

Sukhtankar, S. (2016). India's national rural employment guarantee scheme: What do we really know about the world's largest workfare program?

Wilson, J. and Wildasin, D. (2004). Capital tax competition: bane or boon. *Journal of Public Economics*, 88(6):1065–1091.

Yu, J., Zhou, L.-A., and Zhu, G. (2016). Strategic interaction in political competition: Evidence from spatial effects across Chinese cities. *Regional Science and Urban Economics*, 57:23–37.

Zhang, X., Yang, J., and Wang, S. (2011). China has reached the Lewis turning point. *China Economic Review*, 22(4):542–554.

Zimmermann, L. (2012). Labor market impacts of a large-scale public works program: evidence from the indian employment guarantee scheme. *Working Paper*.

Zimmermann, L. (2014). Public works programs in developing countries have the potential to reduce poverty. *IZA World of Labor*.